



**DEPARTAMENTO DE BIOLOGÍA MOLECULAR**

**OPTIMIZACIÓN DEL SISTEMA REPLICATIVO  
DEL BACTERIÓFAGO  $\phi$ 29 PARA APLICACIONES  
BIOTECNOLÓGICAS**

**TESIS DOCTORAL**

**Pablo Gella Montero**

**Madrid, 2016**



**Departamento de Biología Molecular**  
**Facultad de Ciencias**  
**Universidad Autónoma de Madrid**

**OPTIMIZACIÓN DEL SISTEMA REPLICATIVO  
DEL BACTERIÓFAGO  $\phi$ 29 PARA APLICACIONES  
BIOTECNOLÓGICAS**

Memoria de investigación presentada por el Ldo. en Bioquímica **Pablo Gella Montero** para optar al grado de Doctor en Biología Molecular por la Universidad Autónoma de Madrid.





Esta Tesis Doctoral ha sido realizada en el Centro de Biología Molecular Severo Ochoa (CSIC/UAM) bajo la dirección de la Doctora **Margarita Salas Falgueras** y el Doctor **Mario Mencía Caballero**. Para su realización se contó con una beca predoctoral de la Junta para la Ampliación de Estudios (JAE) del Consejo Superior de Investigaciones Científicas (CSIC).



Abreviaturas.....	1
Glosario de términos no castellanos incluidos en esta Tesis.....	5
Resumen.....	7
<i>Summary</i> .....	9
<b>Introducción .....</b>	<b>11</b>
<b>1. La Biotecnología: origen y avance de una antigua disciplina.....</b>	<b>13</b>
<b>2. La amplificación de ácidos nucleicos .....</b>	<b>14</b>
2.1. La Reacción en Cadena de la Polimerasa (PCR).....	15
2.2. La amplificación isotérmica.....	17
<b>3. El bacteriófago <math>\phi</math>29 .....</b>	<b>18</b>
3.1. Organización genética y transcripcional.....	19
3.2. Los orígenes de replicación .....	20
3.3. La DNA polimerasa del bacteriófago $\phi$ 29 .....	22
3.4. La proteína terminal (TP) del bacteriófago $\phi$ 29.....	24
3.5. La replicación del genoma de $\phi$ 29 .....	25
3.6. Una herramienta biotecnológica .....	29
<b>Objetivos .....</b>	<b>33</b>
<b>Materiales y Métodos.....</b>	<b>37</b>
<b>1. Materiales .....</b>	<b>39</b>
1.1. Nucleótidos y oligonucleótidos.....	39
1.1.1. Nucleótidos.....	39
1.1.2. Oligonucleótidos.....	39
1.2. Plásmidos y construcciones lineales.....	43
1.3. Proteínas y péptidos .....	45
1.4. Estirpes bacterianas .....	45
1.5. Medios de cultivo.....	45

<b>2. Métodos.....</b>	<b>46</b>
2.1. Ensayo de amplificación en presencia de dATP marcado radiactivamente.....	46
2.2. Purificación de oligonucleótidos .....	46
2.3. Hibridación de oligonucleótidos .....	47
2.4. Ensayo de iniciación de la replicación (formación del complejo TP-dAMP).....	47
2.5. Ensayo de transición con oligonucleótidos dsDNA .....	47
2.6. Ensayo de replicación con oligonucleótidos dsDNA .....	48
2.7. Transposición y construcción de la librería de inserciones GFP en la TP de $\phi 29$ .....	48
2.8. Fluorimetría en cultivo líquido .....	49
2.9. Análisis de la expresión de GFP por <i>Western Blot</i> .....	49
2.10. Ensayo de formación del complejo TP-AMP en extractos bacterianos .....	50
2.11. Ensayo de retraso en gel o cambio de movilidad electroforética (EMSA) .....	50
2.12. Ensayo de competición en iniciación con el TP-DNA de $\phi 29$ .....	51
2.13. Ensayo de amplificación detectado por masa.....	51
2.14. Conjugación de la proteína TPi38cys con el grupo maleimida.....	52
 <b>Resultados.....</b>	 <b>53</b>
 <b>Capítulo 1. Método de amplificación de DNAs heterólogos basado en el sistema replicativo del bacteriófago <math>\phi 29</math> .....</b>	 <b>55</b>
1. Diseño de un plásmido cuya linealización produce un DNA flanqueado por los extremos funcionales del genoma de $\phi 29$ .....	57
2. Amplificación <i>in vitro</i> por la maquinaria replicativa del bacteriófago $\phi 29$ del plásmido pETORPHI linealizado .....	59
3. Determinación de la cantidad mínima de molde para la amplificación del plásmido pETORPHI linealizado .....	62
4. Otras estrategias para producir TP-DNAs heterólogos .....	64

**Capítulo 2. Mejora de los orígenes de replicación y estudio de las primeras fases del proceso replicativo del bacteriófago  $\phi 29$  .....69**

1. Efecto de la sustitución y desapareamiento del primer triplete en la eficiencia de utilización del origen de replicación ..... 72
2. Efecto de la sustitución y desapareamiento de los nucleótidos 4-6 en la eficiencia de utilización del origen de replicación ..... 73
3. Efecto de protuberancias en el extremo 5' en la eficiencia de utilización del origen de replicación ..... 76
4. Preferencia del heterodímero TP/DNA polimerasa por la secuencia AAA en el extremo 5' de la banda desplazada..... 79
5. Efecto de la apertura de la hélice en la eficiencia de utilización del origen de replicación ..... 80
6. Pasos afectados en la replicación por los orígenes de replicación desapareados ..... 81
7. Papel del dominio N-terminal de la TP de  $\phi 29$  en la replicación de los orígenes artificiales ..... 83
8. Efecto de la proteína de unión a banda doble p6 en los orígenes de replicación artificiales..... 85

**Capítulo 3. Sustitución del dominio N-terminal de la TP de  $\phi 29$  por el dominio de unión a DNA del factor de transcripción GAL4 ..... 89**

1. Diseño de la proteína TP GAL4 ..... 91
2. Utilización del origen artificial con la secuencia UAS con la proteína TP GAL4 ..... 92
3. Capacidad de unión de la proteína TP GAL4 al DNA ..... 94

**Capítulo 4. Identificación de sitios potenciales de inserción de secuencias peptídicas permitidos en la TP del bacteriófago  $\phi 29$  ..... 97**

1. Aproximación a la identificación de sitios potenciales de inserción permitidos en la secuencia de la TP..... 99
2. Transposición y selección de proteínas ..... 101
3. Iniciación de las TP con insertos y competición por la DNA polimerasa..... 106
4. Capacidad de unión a DNA de las proteínas con insertos..... 108
5. Amplificación del TP-DNA de  $\phi 29$  ..... 110
6. Amplificación de DNA heterólogo: pETORPHI ..... 111

7. La TPi38 y su aplicación biotecnológica .....	112
7.1. Diseño de la aplicación.....	114
7.2. Análisis de la funcionalidad de la TPi38cys .....	114
<b>Discusión .....</b>	<b>117</b>
1. Método de amplificación de DNAs heterólogos basado en el sistema de replicación del bacteriófago $\phi$ 29 .....	119
2. Mejora de los orígenes de replicación y estudio de las primeras fases del proceso replicativo del bacteriófago $\phi$ 29 .....	122
3. Sustitución del dominio N-terminal de la TP de $\phi$ 29 por el dominio de unión a DNA del factor de transcripción GAL4 .....	126
4. Identificación de sitios potenciales de inserción de secuencias peptídicas permitidos en la TP del bacteriófago $\phi$ 29 .....	127
<b>Conclusiones .....</b>	<b>131</b>
<b>Bibliografía .....</b>	<b>135</b>
<b>Material Suplementario.....</b>	<b>155</b>
<b>Artículos científicos publicados .....</b>	<b>159</b>

**Amp<sup>R</sup>**: gen de resistencia a ampicilina.

**ANTP**: péptido antenapedia.

**ARMS**: sistema de mutación refractario a amplificación (del inglés, *Amplification Refractory Mutation System*).

**ATP**: adenosina-5'-trifosfato.

**bdNA**: ácido desoxirribonucleico ramificado (del inglés, *branched DeoxyriboNucleic Acid*).

**BSA**: seroalbúmina bovina (del inglés, *Bovine Serum Albumin*).

**C-terminal**: carboxilo-terminal.

**dAMP**: 2'-desoxiadenosina-5'-monofosfato.

**DAPI**: 4',6-diamino-2-fenilindol.

**dATP**: 2'-desoxiadenosina-5'-trifosfato.

**DBP**: proteína de unión a DNA de banda doble o p6 (del inglés, *Double-stranded Binding Protein*).

**dCTP**: 2'-desoxicitidina-5'-trifosfato.

**DE**: Desviación Estándar.

**dGTP**: 2'-desoxiguanosina-5'-trifosfato.

**DNA**: ácido desoxirribonucleico (del inglés, *DeoxyriboNucleic Acid*).

**dNMP**: 2'-desoxinucleósido-5'-monofosfato.

**dNTP**: 2'-desoxinucleósido-5'-trifosfato.

**DMEM**: modificación de Dulbecco del medio de Eagle (del inglés, *Dulbecco Modification of Eagle Medium*).

**DO<sub>600</sub>**: densidad óptica medida a una longitud de onda de 600 nm.

**ds**: banda doble (del inglés, *double-stranded*).

**DTT**: ditioneitol.

**dTTP**: 2'-desoxitimidina-5'-trifosfato.

**ECL**: quimioluminiscencia electrogenerada (del inglés, *Electrogenerated ChemiLuminescence*).

**EDTA**: ácido etilendiaminotetraacético (del inglés, *EthyleneDiamineTetraacetic Acid*).

**FCS**: suero fetal bovino (del inglés *Fetal Calf Serum*)

**GFP**: proteína verde fluorescente cian (del inglés, *Green Fluorescent Protein*).

**HDA**: amplificación dependiente de helicasa (del inglés, *Helicase-dependent Amplification*).

**HEPES**: ácido 2-[4-(2-Hidroxietil)-1-piperacil]-etanosulfónico.

**IPTG**: isopropil β-D-tiogalactopiranosido.

**ITR**: secuencia terminal repetida inversa (del inglés, *Inverted Terminal Repeat*).

**IVC**: compartimentalización *in vitro* (del inglés, *In Vitro Compartmentalization*).

**Kb**: Kilo base.

**KDa**: Kilo Dalton.

**Km<sup>R</sup>**: gen de resistencia a kanamicina.

**LAMP**: amplificación mediada por bucle (del inglés, *Loop-mediated isothermal Amplification*).

**LB**: medio de cultivo Luria Bertani.

**LCR**: reacción en cadena de la ligasa (del inglés, *Ligase Chain Reaction*).

**MCS**: sitio de policlonaje (del inglés, *MultiCloning Site*).

**MDA**: amplificación por desplazamiento múltiple (del inglés, *Multiple Displacement Amplification*).

**ME**: extremos mosaico del transposón Tn5 (del inglés, *Mosaic Ends*)

**mer**: número de nucleótidos de un oligómero.

**N-terminal**: amino-terminal.

**NASBA**: amplificación de ácidos nucleicos basada en secuencia (del inglés, *Nucleic Acid Sequence Based Amplification*).

**NLS**: secuencia de localización nuclear (del inglés, *Nuclear Localization Signal*).

**OH**: hidroxilo.

**p/v**: peso/volumen.

**PAGE**: electroforesis en gel de poliacrilamida (del inglés, *PolyAcrylamide Gel Electrophoresis*).

**pb**: par de bases.

**PBS**: tampón fosfato salino (del inglés, *Phosphate Buffered Saline*).

**PCR**: reacción en cadena de la polimerasa (del inglés, *Polymerase Chain Reaction*).

**PNK**: polinucleótido kinasa.

**pRNA**: RNA empaquetador (del inglés, *prohead RNA*).

**RBS**: sitio de unión del ribosoma (del inglés, *Ribosome Binding Site*).

**RCA**: amplificación por círculo rodante (del inglés, *Rolling Circle Amplification*).

**RNA**: ácido ribonucleico (del inglés, *RiboNucleic Acid*).

**RNase**: ribonucleasa.

**SDA**: amplificación acoplada a desplazamiento de banda (del inglés, *Strand Displacement Amplification*).

**SDS**: dodecilsulfato sódico (del inglés, *Sodium Dodecyl Sulfate*).

**SH**: tiol.

**ss**: banda sencilla (del inglés, *single-stranded*).

**SSB**: proteína de unión a DNA de banda sencilla o p5 (del inglés, *Single-Stranded Binding protein*).

**sus**: mutante condicional del bacteriófago  $\phi 29$  sensible a la supresión de la mutación que contiene en su genoma (del inglés, *SUPpressor-Sensitive*). Este tipo de fagos son deficientes en la síntesis de alguna proteína viral debido a la introducción en su secuencia de una mutación sin sentido, y sólo pueden replicarse de manera similar al *wild-type* en una estirpe supresora de la mutación. El número que acompaña indica el gen en el que ha sido introducida la mutación.



**TAT:** péptido transactivador de la transcripción del virus de la inmunodeficiencia humana (del inglés, *TransActivator of Transcription*).

**TP:** proteína terminal (del inglés, *Terminal Protein*).

**TP $\Delta$ N:** proteína terminal a la que se le han delecionado los primeros 73 aminoácidos amino terminales.

**TP-dAMP:** complejo covalente formado entre la proteína terminal y el nucleótido iniciador.

**TP-DNA:** DNA con proteína terminal covalentemente unida a los extremos 5'.

**TP-dNMP:** complejo covalente formado entre la proteína terminal y un dNMP.

**TPR:** región de unión a la proteína terminal (del inglés, *Terminal Protein Region*).

**Tris:** tri(hidroximetil) aminometano.

**UDG:** uracil DNA glicosilasa.

**v/v:** volumen/volumen.

**WGA:** amplificación masiva de genomas (del inglés, *Whole Genome Amplification*).

**wt:** tipo natural o silvestre, sin modificación genética (del inglés, *wild-type*).

**XNA:** ácido xenonucleico (del inglés, *XenoNucleic Acid*).

**YFP:** proteína amarilla fluorescente (del inglés, *Yellow Fluorescent Protein*).



**Biopanning:** técnica de selección por afinidad de péptidos que se unen a una diana específica.

**Fingers:** dedos, subdominio del dominio de polimerización de las DNA y RNA polimerasas.

**Footprint:** Técnica de detección de interacciones proteína-DNA basado en el principio de que la unión de la proteína al DNA lo protege de cortes enzimáticos.

**Input:** término empleado para hacer referencia a la cantidad original de DNA de la que se parte en un experimento determinado.

**Jumping-back:** término empleado para describir el mecanismo de translocación hacia atrás del complejo formado por una TP y un trinucleótido tras su síntesis (iniciación de la replicación) para recuperar la información del primer triplete en 3' de la cadena molde.

**Loop:** estructura flexible con forma de bucle o lazo presente en determinados dominios de algunas proteínas.

**Merge:** combinación de imágenes en los experimentos visualizados mediante microscopía de fluorescencia.

**Palm:** palma, subdominio del dominio de polimerización de las DNA y RNA polimerasas.

**Pellet:** material precipitado.

**Sliding-back:** término empleado para describir el mecanismo de translocación hacia atrás del complejo TP-dAMP tras la reacción de iniciación para recuperar la información del primer nucleótido en 3' de la cadena molde.

**Thumb:** pulgar, subdominio del dominio de polimerización de las DNA y RNA polimerasas.

**Western blot:** técnica analítica empleada para la detección de proteínas específicas en una muestra determinada.

**Wild-type:** tipo natural o silvestre, sin modificación genética.



Hace años nuestro laboratorio describió un sistema mínimo para la amplificación eficiente *in vitro* del genoma del bacteriófago  $\phi 29$  basado en 4 proteínas codificadas por el propio virus: la DNA polimerasa, la proteína terminal (TP), la proteína de unión a DNA de banda simple (p5) y la proteína de unión a DNA de banda doble (p6). Sin embargo, hasta el momento no se había desarrollado un sistema similar para la producción de DNAs heterólogos unidos a una TP. En la presente tesis doctoral se ha adaptado el sistema de amplificación cebado por TP a DNAs que no tienen una TP unida inicialmente a sus extremos 5'. Esto se ha realizado mediante el diseño de distintos abordajes que permiten flanquear los DNAs a amplificar con las secuencias terminales del genoma de  $\phi 29$ , y se ha demostrado la posibilidad de obtener con ellos una amplificación eficiente *in vitro* cuyo producto es un DNA heterólogo unido covalentemente a una TP en cada uno de sus extremos 5'.

Con el fin de comprender los eventos tempranos y para encontrar orígenes de replicación mejorados para la amplificación cebada por TP, en la presente tesis doctoral se ha realizado un estudio sistemático de diferentes secuencias y estructuras terminales en orígenes artificiales de dsDNA. Se observó que la eficiencia del origen estaba determinada por una combinación de factores, entre ellos la secuencia y el desapareamiento de la doble banda y, además, era necesario el reconocimiento de una estructura tipo 'origen' (dsDNA y romo). El origen más eficiente obtenido (30 veces la actividad *wild-type*) tenía la secuencia CCC en el extremo 3' de la hebra molde, AAA en el extremo 5' de la hebra desplazada y los 6 primeros pares de bases desapareados. El incremento en la actividad de los orígenes mejorados parecía ser el resultado de la reducción de una barrera termodinámica en la etapa de transición.

Trabajos anteriores han demostrado que el dominio N-terminal de la TP presenta capacidad de unión a DNA de forma inespecífica de secuencia. En la presente tesis doctoral se propuso generar una unión específica como forma de mejorar el acceso de la maquinaria de replicación al origen y, por lo tanto, la eficiencia de replicación. Para ello, se sustituyó el dominio N-terminal de la TP por el dominio de unión a DNA del factor de transcripción GAL4 y se generó una diana de reconocimiento para GAL4 haciendo los mínimos cambios posibles en un origen basado en la secuencia del extremo derecho de  $\phi 29$ . Sin embargo, la proteína no resultó funcional, habiendo perdido la capacidad de unión a DNA.

La unión del genoma de  $\phi 29$  a la TP, codificada por él mismo, hace del sistema del bacteriófago  $\phi 29$  una plataforma muy interesante para fusionar nuevas funciones a un DNA y desarrollar aplicaciones tipo DNA display. En la presente tesis doctoral, se ha empleado la transposición al azar de GFP en la secuencia de la TP con el objeto de encontrar sitios potenciales de inserción seleccionando 12 proteínas que fueron analizadas con el sistema replicativo de  $\phi 29$ . De éstas, una permite una inserción de hasta 17 aminoácidos en su secuencia, manteniendo la funcionalidad de la proteína, y puede ser la base de nuevas tecnologías que aún están en proceso de desarrollo, concretamente para la transferencia de genes.



Some years ago our laboratory described a minimal system for the *in vitro* efficient amplification of the genome of bacteriophage  $\phi 29$ . This minimal system is based in 4 phage-encoded proteins: DNA polymerase (p2), Terminal Protein (TP), single-stranded DNA binding protein (p5) and double-stranded DNA binding protein (p6). Until now, this amplification system has not been developed to produce defined TP-linked heterologous DNAs in a fashion similar to the amplification of the phage  $\phi 29$  genome. In the present work we adapt this system to DNAs lacking a TP in their 5' ends. For that, we have designed different approaches that allow the flanking of these ends with the terminal sequences of the genome of  $\phi 29$ , demonstrating the possibility of obtaining an *in vitro* efficient amplification whose product is an heterologous DNA covalently linked to a TP at both 5' ends.

To understand the early replication events and find improved replication origins for the TP-primed amplification, we systematically analyzed different sequences and terminal structures in dsDNA artificial origins. We observed that origin efficiency depends on several factors like the sequence and the double strand unpairing, and also that a structurally recognizable origin was needed (dsDNA and blunt). The strongest origin (30-fold *wild-type* activity) had the sequence CCC at the 3' end, AAA at the 5' end and its 6 first base pairs unpaired. The increased activity of the origins seemed to be the result of the reduction of a thermodynamic barrier at the transition step.

Previous works have shown that the TP N-terminal domain possesses non-specific DNA binding capacity *in vitro*. In this work, we proposed generating a specific DNA binding to improve the access of the replication machinery to the origin and, therefore, the replication efficiency. For this, we replaced the TP N-terminal domain with the DNA binding domain of the GAL4 transcription factor. We also generated the recognition target for GAL4 on an origin based in the right  $\phi 29$  DNA end sequence making the minimum number of changes possible. However, the protein was not functional, having lost the DNA binding ability.

The linkage of the  $\phi 29$  genome to the TP, encoded by itself, makes the bacteriophage  $\phi 29$  system an interesting platform to fuse new functions to a DNA or to develop applications such as DNA display. In this work, we used the random GFP transposition into the TP sequence to find potencial insertion sites. We selected 12 proteins that were analyzed with the replicative system of  $\phi 29$ . One of these proteins has proved to be active even with a 17 amino acid insertion, and it could be the basis for new technologies that are still under development, namely gene transfer.





# Introducción



## 1. La Biotecnología: origen y avance de una antigua disciplina

Poco podía imaginar Karl Ereky el alcance del término que emplearía por primera vez en 1919 para referirse a la intersección entre biología y tecnología humana: *Biotecnología* (Ereky, 1919). En su trabajo Ereky proponía un sistema de cría de cerdo a gran escala basado en su alimentación con remolacha azucarera, lo que ponía de manifiesto que, si bien la biotecnología era un sector neonato, su práctica ya se daba desde el surgimiento de la agricultura en la Revolución del Neolítico. Lo que en su día fue un simple nombre, en la actualidad es sinónimo de modernidad, haciendo que muchos especialistas y diletantes hayan calificado al siglo XXI como el siglo de la biotecnología.

Por otra parte, apenas han pasado 80 años desde que el matemático Warren Weaver acuñara, en un contexto de importantes hitos científicos, el término *Biología Molecular* en 1938 para designar a la aplicación de los conocimientos y herramientas de las Ciencias Físicas al estudio de la materia viva. Su nacimiento trajo consigo una serie de avances concentrados en un período muy corto de tiempo que supusieron una auténtica revolución científica, aunque la irrupción de la II Guerra Mundial (1939-1945) relegó a un segundo plano la importancia de trabajos que posteriormente se convertirían en sus cimientos. Algunos de estos trabajos incluyen los de Beadle y Tatum, que, bajo la hipótesis “un gen, un enzima”, demostraron la existencia de una correlación entre genes y enzimas (Beadle y Tatum, 1941), o los de Avery, MacLeod y McCarty, que aportaron la primera evidencia científica de que el material genético era el DNA (Avery *et al.*, 1944). Sin embargo, lo que para muchos significó el mayor hito en la Biología Molecular fue la propuesta del modelo estructural del DNA de Watson y Crick (Watson y Crick, 1953), que integraba los datos químicos sobre su composición molar de bases (Chargaff, 1950) y los datos físicos obtenidos por difracción de rayos X (Franklin y Gosling, 1953). Los profundos conocimientos que se fueron adquiriendo a partir de entonces sobre los ácidos nucleicos culminarían en la década de los 70 en el desarrollo de un conjunto de técnicas que permitían su manipulación bajo el nombre de *Ingeniería Genética*, y que comenzaría con el descubrimiento, uso y aplicación del primer enzima de restricción (Arber, 1965; Smith y Wilcox, 1970; Danna y Nathans, 1971). Los enormes avances científicos del siglo XX influyeron decisivamente en la biotecnología e impusieron la discriminación entre una biotecnología tradicional y otra moderna. Esta última emplearía las nuevas técnicas de la ingeniería genética y se definiría como la aplicación tecnológica que usa sistemas biológicos o derivados para la creación o modificación de productos o procesos comerciales o aplicables. De esa forma, más que una ciencia en sí misma, la biotecnología se entiende como un área multidisciplinar que integra disciplinas científicas de conocimiento básico y aplicado con el objetivo de producir beneficios para el ser humano. En realidad, esto es el resultado de la propia evolución del pensamiento científico que, si bien en un principio se centraba en analizar sus objetos de estudio como elementos y procesos aislados, hoy en día lo hace en la búsqueda de su aplicación. Esta evolución también ha

traído consigo un cambio estructural en la composición del grupo investigador, de forma que el individuo aislado ha ido dando paso a los modernos equipos multidisciplinares que operan de forma colaborativa. Este carácter integrador también se observa en el aspecto social, que busca ser informado y se muestra en la cada vez mayor concienciación y preocupación por los avances en el sector. De esta forma, los enormes progresos en las técnicas se suceden de forma acelerada. La gran cantidad de información disponible y el enfoque multidisciplinar ha permitido que modelos de estudio, de los que ya se cuenta con un amplio conocimiento, se aborden con nuevos métodos o adaptaciones tecnológicas y se sitúen en un nuevo plano, como es el caso del protagonista de la presente tesis doctoral: el bacteriófago  $\phi 29$ .

## 2. La amplificación de ácidos nucleicos

La amplificación de ácidos nucleicos es un proceso fundamental en biotecnología y biología molecular y su uso se extiende por los diversos campos de especialización. Entre sus aplicaciones más destacadas se encuentra su uso en diagnóstico molecular, secuenciación, mutagénesis dirigida, clonación y creación de librerías genéticas. Particularmente, gracias a su desarrollo se ha conseguido descifrar el mapa genético de la especie humana en el Proyecto Genoma Humano. Además, el conocimiento del mapa de DNA específico de cada individuo ha facilitado los primeros pasos en el desarrollo de la denominada medicina personalizada.

En general, las técnicas de amplificación han de cumplir ciertos requisitos, incluyendo sensibilidad, grado de automatización y coste reducido, que permitan maximizar su eficiencia y aplicabilidad. La evolución y diversificación en función de su aplicación y tecnología empleada ha permitido la elaboración de una clasificación para estas técnicas (Tang y Persing, 2009):

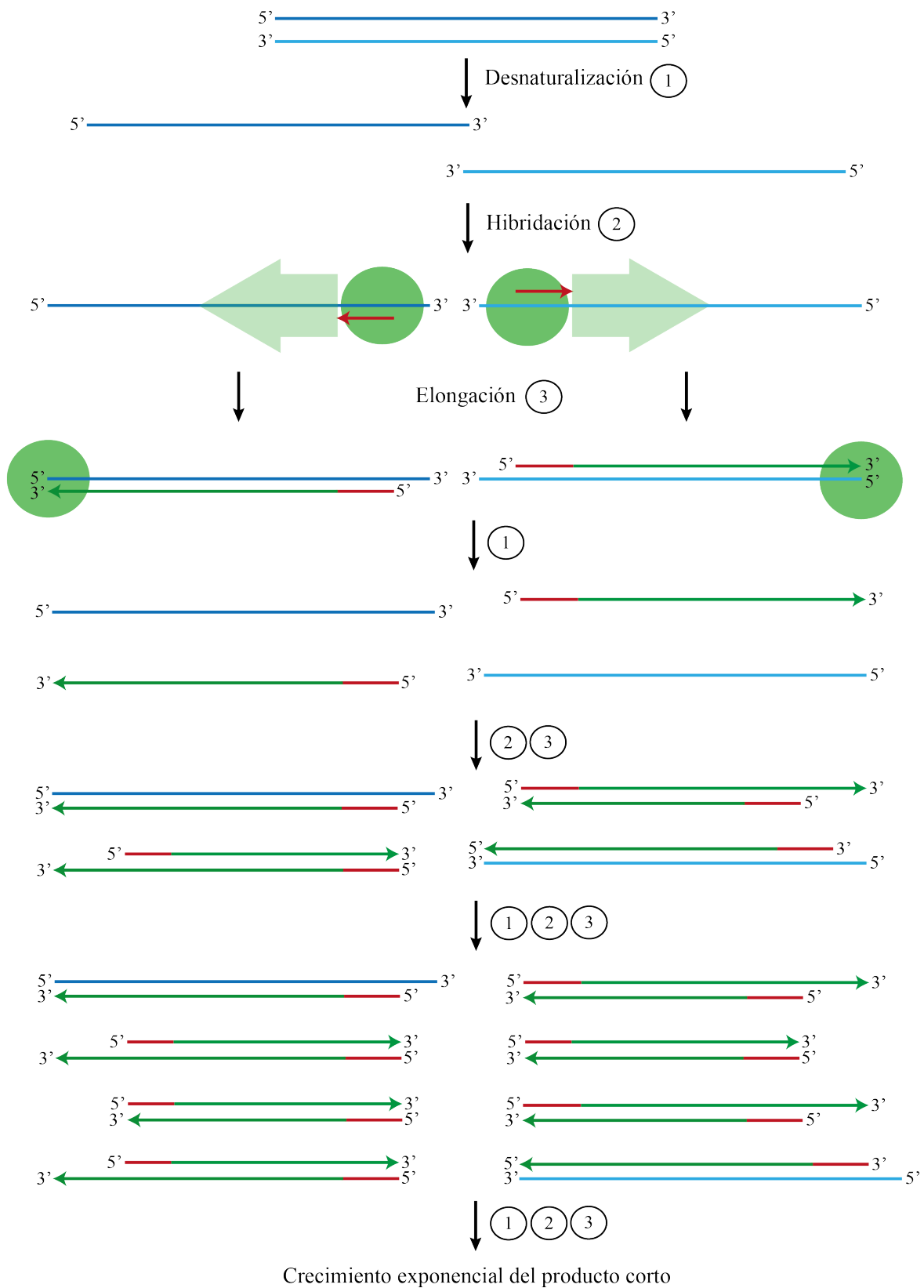
(I) Técnicas de amplificación de diana: Son técnicas basadas en la amplificación enzimática de una molécula diana hasta el nivel de detección, por lo que son útiles en situaciones donde la secuencia a amplificar es importante y/o se requiere gran cantidad de ella para futuros usos. Esta categoría incluye, entre otras, las técnicas de la Reacción en Cadena de la Polimerasa (PCR) (Saiki *et al.*, 1985), la Amplificación de Ácidos Nucleicos Basada en Secuencia (NASBA) (Guatelli *et al.*, 1990; Compton *et al.*, 1991) y la Amplificación por Desplazamiento Múltiple (MDA) (Dean *et al.*, 2002).

(II) Técnicas de amplificación de sonda: Son técnicas basadas en la amplificación de las sondas que reconocen específicamente una diana de interés, por lo que son útiles cuando la información de secuencia de la diana no es importante pero sí su presencia. Esta categoría incluye, entre otras, las técnicas de Reacción en Cadena de la Ligasa (LCR) (Wu y Wallace, 1989; Barany, 1991) y la Amplificación de sondas de RNA con la Q $\beta$  replicasa (Lizardi *et al.*, 1988; Pritchard *et al.*, 1990).

(III) Técnicas de amplificación de señal: Son técnicas empleadas para incrementar la sensibilidad de las basadas en hibridación de sonda, por lo que son útiles cuando la señal esperada es muy baja o el material de partida es escaso. Esta categoría incluye, entre otras, la técnica de las Sondas ramificadas de DNA (bDNA) (Collins *et al.*, 1997)

### 2.1. La Reacción en Cadena de la Polimerasa (PCR)

Posiblemente, cualquier biólogo molecular apuntaría a la técnica de la Reacción en Cadena de la Polimerasa (en inglés, PCR) como uno de los hitos más importantes y con mayor trascendencia en la historia de las Ciencias de la Vida. La PCR es una técnica de amplificación de diana basada en la amplificación exponencial de un segmento específico de DNA partiendo de mínimas cantidades de muestra. Para ello, se utilizan dos oligonucleótidos denominados cebadores que tienen una secuencia complementaria al inicio y final del fragmento a amplificar y una mezcla de reacción que contiene el DNA molde, dNTPs libres, sales y una DNA polimerasa. La PCR es un proceso simétrico donde el desplazamiento de la banda no copiada se hace por medio de la desnaturalización térmica (**Figura 1**). Así, un termociclador se encarga de repetir ciclos de temperatura que incluyen los siguientes pasos: (i) *Desnaturalización* de las hebras del DNA, (ii) *Hibridación* de los cebadores a las secuencias flanqueantes y (iii) *Elongación* de los cebadores mediante una copia fiel por la DNA polimerasa de la hebra de DNA a la que se han hibridado. De esta forma, las moléculas de DNA de nueva síntesis sirven como moldes en los ciclos posteriores, produciendo una amplificación exponencial de un billón de copias en 30 ciclos (Tenover *et al.*, 1997; Saunders *et al.*, 2003). Aunque hoy en día su uso extensivo y accesibilidad difuminan su relevancia, la técnica posee una larga historia a sus espaldas escrita por las demandas del tirón científico del último siglo. Lo que en un principio comenzó siendo el estudio del mecanismo de reparación por las DNA polimerasas replicativas (Kleppe *et al.*, 1971) constituyó la prueba de concepto para lo que más tarde conduciría a Kary Mullis a desarrollar una técnica de amplificación de DNA, en colaboración con otros científicos de la compañía Cetus (Saiki *et al.*, 1985), usando el fragmento Klenow de la DNA polimerasa I de *Escherichia coli*. Sin embargo, las etapas de alta temperatura dentro de los ciclos inactivaban dicho enzima haciendo necesario un recebado con polimerasa al final de cada ciclo. La sustitución del fragmento Klenow por la DNA polimerasa termoestable de *Thermus aquaticus* permitió una reacción continua y una mayor fidelidad de copia, ya que el aumento de temperatura evitaba los productos inespecíficos (Saiki *et al.*, 1988; Lawyer *et al.*, 1993). A lo largo del tiempo, la técnica ha sido mejorada, permitiendo en la actualidad la producción de gran cantidad de DNA en un período muy corto de tiempo (Ahmed *et al.*, 2012), lo que ha sido la base de numerosas aplicaciones en diagnóstico clínico y Biología Molecular. Además, en los casos donde la amplificación por la técnica de PCR básica ha resultado poco exitosa se han desarrollado modificaciones para hacerla más versátil. Este es el caso



**Figura 1. Esquema del proceso de amplificación de DNA por PCR.** Representación esquemática del proceso de la Reacción en Cadena de la Polimerasa (PCR). Se muestran los pasos de *desnaturalización* (1), *hibridación* (2) y *elongación* (3).

de la Multiplex PCR (Chamberlain *et al.*, 1988), que permite la coamplificación de diferentes dianas en la misma reacción; la PCR anidada (Snounou *et al.*, 1993) que permite un incremento sustancial en la especificidad del proceso; o el Sistema de Mutación Refractario a Amplificación (ARMS) (Newton *et al.*, 1989), que permite la identificación de mutaciones específicas en la secuencia diana.

## 2.2. La amplificación isotérmica

A pesar de su importancia, la PCR tiene una serie de limitaciones inherentes a la propia técnica como la necesidad de un termociclador y una eficiencia y especificidad que dependen exclusivamente de los primeros ciclos. Además, el diseño de cebadores implica la necesidad de disponer de una secuencia conocida, lo cual muchas veces no es posible como en el caso de los metagenomas, y el rendimiento de la técnica disminuye cuando se emplean cebadores largos o temperaturas de anillamiento muy dispares dentro del grupo de cebadores utilizado. Por otra parte, las limitaciones de las DNA polimerasas empleadas hacen que la amplificación de fragmentos de DNA o genomas muy grandes sea muy ineficiente de forma que parte de la secuencia diana no es cubierta de manera representativa. A esto se le unen otros problemas como el elevado riesgo de contaminación y la sensibilidad a ciertos compuestos o inhibidores. Por todo esto, se han desarrollado métodos alternativos de amplificación de ácidos nucleicos, gran parte de los cuales se agrupa bajo las conocidas como técnicas de amplificación isotérmica. Las más conocidas incluyen la amplificación mediada por bucle (LAMP), la amplificación acoplada a desplazamiento de banda (SDA), la amplificación dependiente de helicasa (HDA) y las técnicas de amplificación por círculo rodante (RCA) y amplificación por desplazamiento múltiple (MDA). Estas dos últimas se basan en dos características únicas contenidas en la DNA polimerasa del bacteriófago  $\phi 29$ : la capacidad de desplazamiento de banda y la alta procesividad.

La amplificación isotérmica mediada por bucle (LAMP) emplea 4-6 cebadores que reconocen varias regiones del DNA diana, dos de los cuales forman estructuras en forma de bucle que facilitan posteriores rondas de amplificación gracias a una DNA polimerasa con capacidad de desplazamiento de banda (Notomi *et al.*, 2000).

La amplificación acoplada a desplazamiento de banda (SDA) emplea un enzima de restricción que corta en una de las hebras para producir una mella y una DNA polimerasa con capacidad de desplazamiento de banda, comúnmente la BstDNA polimerasa, para regenerar el corte y elongar la hebra a partir de dicha posición (Walker *et al.*, 1992).

La amplificación dependiente de helicasa (HDA) emplea la capacidad de una helicasa para separar las hebras del DNA, eliminando el paso de desnaturalización térmica, permitiendo la hibridación de los oligonucleótidos y su extensión por una DNA polimerasa (Vincent y Kong, 2004).

La amplificación por círculo rodante (RCA) se basa en la elongación de un extremo 3'OH proporcionado por un oligonucleótido cebador o por la producción de una mella en el DNA molde para la producción de ssDNA a partir de un molde de DNA circular. De esta forma, el producto de la amplificación es un concatémero que contiene múltiples copias en tándem complementarias al DNA molde que pueden ser procesadas posteriormente (Dean *et al.*, 2001).

La amplificación por desplazamiento múltiple (MDA) emplea hexámeros con una secuencia aleatoria que hibridan en diferentes partes del DNA a amplificar. Estos son elongados por una DNA polimerasa con capacidad de desplazamiento de banda, lo que genera cadenas de DNA de banda simple laterales susceptibles de hibridar más hexámeros. De esa forma, la amplificación resulta en la formación de una red altamente ramificada de DNA (Dean *et al.*, 2002). La MDA ha sido la principal técnica empleada en la secuenciación masiva de genomas (WGA) gracias a las dos características mencionadas anteriormente de la DNA polimerasa de  $\phi 29$  y a su capacidad correctora de errores. Mientras que otras polimerasas empleadas en WGA presentaban un tasa relativamente alta de mutación (tasa de error de  $10^{-5}$ ) (Silander y Saarela, 2008), la DNA polimerasa de  $\phi 29$  tiene una frecuencia de error de  $10^{-6}$ - $10^{-7}$  (Esteban *et al.*, 1993). Esto es comparable a la del enzima *Pfu* ( $7 \times 10^{-7}$ ) u otras DNA polimerasas con actividad correctora de errores (Cha *et al.*, 1995), siendo significativamente menor a la de la *Taq* ( $2 \times 10^{-4}$ ) (Dunning *et al.*, 1988).

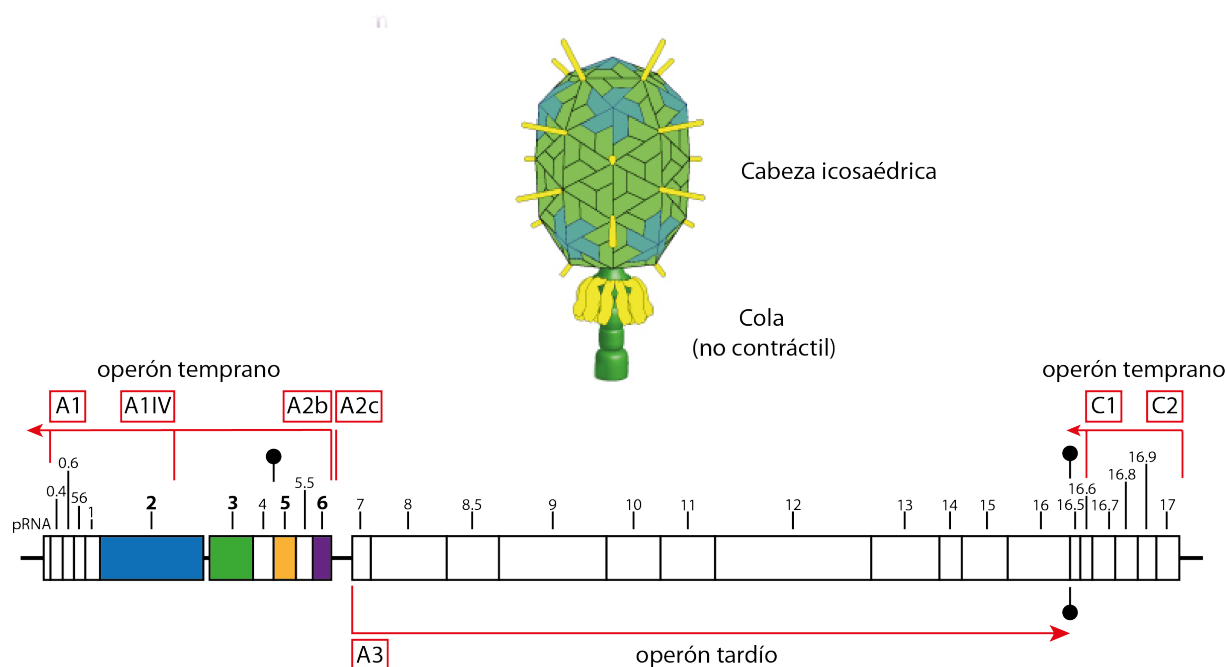
Algunos aspectos siguen constituyendo un problema como los DNAs con un alto contenido en G/C o con secuencias repetidas, bien por la especificidad en la hibridación de los cebadores o por la formación de estructuras secundarias. Estas estructuras pueden conducir a un deslizamiento de la polimerasa con la pérdida de la información correspondiente (Viguera *et al.*, 2001), o una disociación de la misma, resultando productos de amplificación no específicos o abortivos y una incompleta cobertura del replicón (Dean *et al.*, 2002).

### 3. El bacteriófago $\phi 29$

Elegido inicialmente por la condición de ser pequeño pero suficientemente complejo desde un punto de vista estructural y poco conocido, el bacteriófago  $\phi 29$  se ha consolidado desde su descubrimiento (Reilly y Spizizen, 1965) y a lo largo de los años como un modelo de investigación básica en Biología Molecular. Fundamentalmente, su estudio ha proporcionado gran cantidad de información sobre los procesos básicos de replicación y transcripción del DNA, así como de las interacciones virus-hospedador. El bacteriófago  $\phi 29$  es un virus lítico de la familia *Podoviridae* (cabeza icosaédrica y cola corta no contráctil) cuyo hospedador natural es la bacteria Gram-positiva *Bacillus subtilis*, aunque puede infectar a otras bacterias del género *Bacillus* (Ackermann, 1998). El género de fagos  $\phi 29$ -like agrupa, en base a similitud de secuencia, a los fagos  $\phi 29$ , PZA,  $\phi 15$  y BS32 (grupo I); B103, Nf y M2Y (grupo II); y GA-1 (grupo



III). Estos virus son muy similares, además de en la secuencia genética, en sus proteínas y modo en que replican su genoma (Pečenková y Pačes, 1999).



**Figura 2. Mapa genético y transcripcional del bacteriófago  $\phi 29$ .** En la parte superior se muestra el modelo morfológico del virión de  $\phi 29$ , indicando las dos principales características de los bacteriófagos pertenecientes a la familia *Podoviridae* (cabeza icosaédrica y cola corta no contráctil). En la parte inferior se muestra el mapa genético y transcripcional de  $\phi 29$ . Se indica en ambos extremos del genoma los operones de expresión temprana izquierdo y derecho, así como el operón de expresión tardía entre ellos. Se indican los promotores más importantes en cajas rojas y los terminadores de transcripción mediante horquillas. Las flechas rojas indican la dirección de la transcripción. De todos los genes, se muestran coloreados los que corresponden a las principales proteínas implicadas en la replicación *in vitro* del genoma de  $\phi 29$ , de gran relevancia para la presente tesis doctoral (adaptado de Xiang *et al.*, 2008).

### 3.1. Organización genética y transcripcional

El genoma del bacteriófago  $\phi 29$  consta de un DNA lineal de doble banda (dsDNA) de 19285 pb (Vlček y Pačes, 1986) cuyos extremos 5' están covalentemente unidos a sendas moléculas de proteína terminal o TP (Salas *et al.*, 1978). A esta TP unida al DNA se le denomina TP parental para diferenciarla de la TP libre de nueva síntesis.

La transcripción del genoma es llevada a cabo por la RNA polimerasa bacteriana y se estructura en 3 operones transcripcionales: dos tempranos situados a ambos extremos del DNA y transcritos de derecha

a izquierda, y uno tardío que ocupa la parte central y es transcrito en el sentido contrario (Sogo *et al.*, 1979) (**Figura 2**).

El operón temprano derecho es el primero que se transcribe (González-Huici *et al.*, 2004a) y codifica proteínas implicadas en la inyección del material genético y la replicación en el interior de la bacteria (González-Huici *et al.*, 2004a, 2006; Alcorlo *et al.*, 2007; Meijer *et al.*, 2000; Meijer *et al.*, 2001b; Serna-Rico *et al.*, 2002, Serna-Rico *et al.*, 2003).

El operón temprano izquierdo codifica las principales proteínas implicadas en la replicación del genoma de  $\phi 29$  (DNA polimerasa, TP, p5 y p6). La DNA polimerasa y la TP son codificadas por los genes 2 y 3, respectivamente (Peñalva y Salas, 1982; Blanco y Salas, 1984). La proteína de unión a DNA de banda simple (SSB) o p5 y la proteína de unión a DNA de banda doble (DBP) o p6 son codificadas por los genes 5 y 6, respectivamente (Martín *et al.*, 1989; Serrano *et al.*, 1993). Este operón también codifica la proteína p56, un inhibidor de la uracil-DNA glicosilasa (UDG) bacteriana (Serrano-Heras *et al.*, 2008; Pérez-Lago *et al.*, 2011; Asensio *et al.*, 2011); la proteína de unión a membrana p1 (Bravo y Salas, 1997), cuya expresión produce un incremento en la longitud celular y número de partículas virales (Ballesteros-Plaza *et al.*, 2013); la proteína de regulación transcripcional p4, que regula el paso de la transcripción temprana a la tardía (Sogo *et al.*, 1979); y la síntesis del pRNA, que interviene en la morfogénesis y el mecanismo de encapsidación (Zhang *et al.*, 1994; Garver y Guo, 1997).

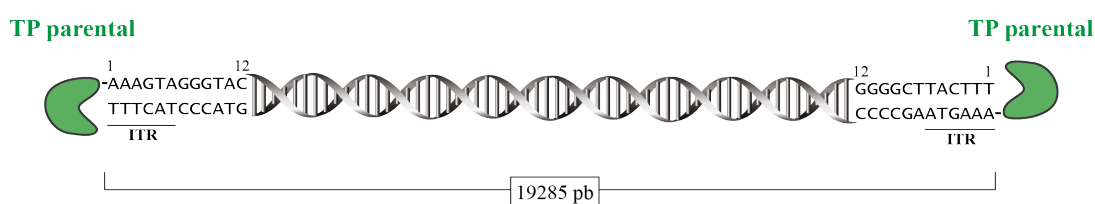
El operón tardío codifica las proteínas estructurales, así como las implicadas en la morfogénesis de la partícula viral y en la lisis celular (revisado en Meijer *et al.*, 2001a; Steiner *et al.*, 1993).

### 3.2. Los orígenes de replicación

La incapacidad de las DNA polimerasas para comenzar *de novo* la síntesis de DNA y su sentido exclusivo de polimerización 5'-3' ha exigido el desarrollo de estrategias para solucionar el problema derivado de pérdida de información genética en los DNA lineales. Ejemplos de ello son los bacteriófagos T4, T7 y SPP1, que forman concatémeros durante la replicación; el bacteriófago  $\lambda$ , que recirculariza su genoma (revisado en Salas y de Vega, 2008); o la actividad del enzima telomerasa para evitar el progresivo acortamiento de los extremos cromosomales en eucariotas superiores (Verdun and Karlseder, 2007).

Por otra parte, algunos bacteriófagos como  $\phi 29$ , Nf, GA-1, PRD1, Cp-1 o el virus en forma de botella de *Acidianus* (Bamford *et al.*, 1995; García *et al.*, 1997; Meijer *et al.*, 2001; Peng *et al.*, 2007), plásmidos lineales de bacterias como pSCL y pSLA2 (Chang *et al.*, 1996), adenovirus (Liu *et al.*, 2003), plásmidos lineales de mitocondria, levaduras, plantas y cromosomas de algunas cepas bacterianas (*Streptomyces* sp.) (revisado en Salas, 1991; Hinnebusch y Tilly, 1993; Kapitonov y Jurka, 2006; Fricova *et al.*, 2010), han desarrollado un sistema replicativo cebado por TP que mimetiza los cebadores de RNA o DNA

en los extremos del genoma, evitando la pérdida de información (Salas, 1999). Estas TPs proporcionan el grupo hidroxilo libre (-OH) necesario para iniciar la replicación del DNA y requieren generalmente de DNA polimerasas especiales (Salas, 1991). Generalmente, estas TPs utilizan unas secuencias determinadas en el extremo del DNA y, en algunos casos, proteínas accesorias, de forma que los extremos de estos genomas con TP constituyen auténticos orígenes de replicación donde existe una compleja red de interacciones proteína-proteína y proteína-DNA. Aunque esta organización genómica es más típica de fagos



**Figura 3. Orígenes de replicación del bacteriófago  $\phi 29$ .** El genoma de  $\phi 29$  es un DNA lineal de doble cadena y 19285 pb. Los orígenes mínimos de replicación se localizan en los extremos del DNA y constan de una proteína terminal (TP) unida covalentemente a los extremos 5' y una secuencia comprendida en los 12 primeros pb. Estas posiciones contienen las ITR, de 6 pb, y éstas, a su vez, el triplete TTT necesario para una eficiente iniciación de la replicación.

o plásmidos (Salas, 1999), se han encontrado elementos de DNA lineal con proteínas unidas de forma covalente a sus extremos 5' en los tres dominios de la vida, como se ha mencionado más arriba. De todos, el sistema replicativo de  $\phi 29$  ha sido el más extensamente estudiado (Salas y de Vega, 2006).

En la mayor parte de los casos, estos sistemas contienen en los extremos del genoma secuencias terminales repetidas inversas (ITR) (Salas, 1991) que se han propuesto como secuencias de reconocimiento por parte del complejo formado por la TP y la DNA polimerasa libre, así como por otras proteínas accesorias de unión a DNA que intervienen en el proceso de la iniciación de la replicación cebada por proteína terminal (Salas, 1999; de Jong *et al.*, 2003; Bamford *et al.*, 1995; Bao y Cohen, 2003). En el caso de Adenovirus, estas ITRs constan de aproximadamente 100 pb, dependiendo del serotipo, y son reconocidas por la TP, la DNA polimerasa y los factores celulares NFI y NFII para su óptima eficiencia (de Rong *et al.*, 2003). En el caso de los plásmidos lineales o cromosomas del género *Streptomyces*, se ha observado que las ITR unen a la TP y a una proteína de unión a DNA denominada Tap (Bao y Cohen, 2003; Yang *et al.*, 2002; Huang *et al.*, 2007). En algunas ocasiones, éste no siempre es el único mecanismo de replicación de DNA lineal. En el caso de algunos plásmidos de *Streptomyces* existen orígenes de replicación circular desde los que se replica la mayor parte del DNA, quedando el mecanismo de replicación cebado por TP

como uno de mantenimiento de extremos similar al telomérico (Shiffman y Cohen, 1992; Chang y Cohen, 1994).

En el caso particular de  $\phi 29$ , aunque estudios de delección y mutagénesis puntual mostraron que unos posibles orígenes mínimos de replicación estaban comprendidos en los 12 primeros pares de bases (pb) (Gutiérrez *et al.*, 1988), su secuencia no ha sido definida con exactitud. Estas posiciones contenían las ITR, de 6 pb y secuencia 5'AAAGTA' (Escarmís y Salas, 1981; Yoshikawa *et al.*, 1981), y éstas, a su vez, el triplete TTT necesario para la iniciación eficiente de la replicación (**Figura 3**). Por otra parte, aunque la presencia de la TP paterna es esencial para la replicación *in vivo* del DNA (Escarmís *et al.*, 1989), su eliminación total con piperidina resulta en un DNA molde funcional para su replicación *in vitro* aunque con una eficiencia 30 veces menor (García *et al.*, 1984; Gutiérrez *et al.*, 1986). Además, ensayos de iniciación con sistemas mixtos (DNA polimerasa, TP y TP-DNA) entre  $\phi 29$  y Nf mostraron actividades mayores cuando la DNA polimerasa y el TP-DNA eran del mismo fago, lo que sugería un reconocimiento específico de la TP paterna por parte de la DNA polimerasa (González-Huici *et al.*, 2000b). Por su parte, el análisis de una DNA polimerasa quimérica de  $\phi 29$  a la que se le había sustituido el subdominio TPR1 por el del fago GA-1 mostró que empleando la TP iniciadora de GA-1, únicamente era capaz de llevar a cabo la reacción de iniciación en presencia del TP-DNA de  $\phi 29$  como molde, lo que indicaba que la mayor contribución en el reconocimiento de los orígenes de replicación recaía sobre la propia DNA polimerasa (Pérez-Arnáiz *et al.*, 2007). Sin embargo, análisis de mutantes de la TP de  $\phi 29$  indicaron que los residuos Asn80, Tyr82 y Arg169 también estaban implicados en la identificación de la TP paterna (del Prado *et al.*, 2012; Illana *et al.*, 1999). Además del reconocimiento de los orígenes de replicación por el heterodímero, la proteína p6 y su unión en los extremos del genoma se han propuesto como un mecanismo de reclutamiento de la maquinaria de replicación (complejo DNA polimerasa/TP) al origen (Freire *et al.*, 1996; Pérez-Arnaiz *et al.*, 2007).

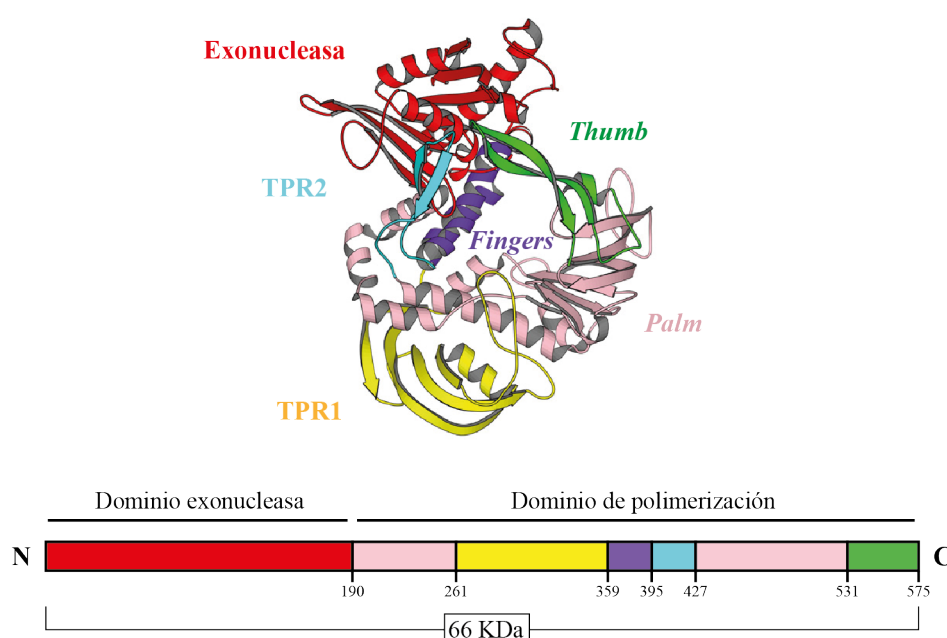
### 3.3. La DNA polimerasa del bacteriófago $\phi 29$

La alta procesividad (síntesis de >70 Kb sin disociación del DNA) y la capacidad de polimerización acoplada a desplazamiento de banda (Blanco y Salas, 1985b; Blanco *et al.*, 1989), ambas características contenidas en el mismo enzima, han convertido a la DNA polimerasa del bacteriófago  $\phi 29$  en un paradigma de la aplicación biotecnológica a partir de la investigación básica. Esto, unido a la capacidad correctora de errores ha sido la base para su uso en un número cada vez mayor de métodos de amplificación isotérmica de DNA.

La DNA polimerasa de  $\phi 29$  es un enzima monomérico de 66 KDa codificada por el gen 2 del genoma (Blanco *et al.*, 1984) que pertenece a la familia B de las DNA polimerasas dependientes de DNA y,

concretamente, al subgrupo de polimerasas que inician la replicación con TP (Bernad *et al.*, 1987; Blanco *et al.*, 1986).

Funcionalmente, la DNA polimerasa presenta dos dominios bien diferenciados (Blanco y Salas, 1995; Blanco y Salas, 2006; Kamtekar *et al.*, 2004): un dominio de polimerización (C-terminal) que agrupa las actividades polimerasa (Blanco *et al.*, 1984) y pirofosforolítica (Blasco *et al.*, 1991); y un dominio exonucleasa (N-terminal) con una actividad exonucleasa 3'-5' correctora de errores (Blanco y Salas, 1985a) (**Figura 4**). Su actividad de polimerización ostenta unos valores de discriminación en la inserción del nucleótido de  $10^4$ - $10^6$  y elonga desapareamientos  $10^5$ - $10^6$  veces peor que pares correctos (Esteban *et al.*, 1993).



**Figura 4. Estructura tridimensional y mapa funcional de la DNA polimerasa de  $\phi 29$ .** La DNA polimerasa de  $\phi 29$  es un enzima monomérico de 66 KDa con dos dominios funcionales bien diferenciados: un dominio exonucleasa (N-terminal) y un dominio de polimerización (C-terminal). Se representan con colores los subdominios exonucleasa (rojo), *palm* (rosa), TPR1 (amarillo), *fingers* (morado), TPR2 (azul) y *thumb* (verde). El modelado se realizó con el software Pymol (<http://www.pymol.org>) a partir de los datos cristalográficos contenidos en el número de acceso **1XII** de la Protein Data Bank (PDB).

Gracias a la resolución de la estructura cristalográfica se observó que, además de los tres subdominios característicos de las DNA polimerasas, *palm*, *fingers* y *thumb*, descritos por primera vez en la DNA polimerasa I de *E. coli* (Ollis *et al.*, 1985), la DNA polimerasa de  $\phi 29$  cuenta con dos regiones adicionales.

les conservadas en aquellas que inician la replicación con TP, denominadas *Terminal Protein Region 1* (TPR1) y *Terminal Protein Region 2* (TPR2) (Blasco *et al.*, 1990; Dufour *et al.*, 2000) (**Figura 4**). El subdominio TPR1, localizado entre *palm* y *fingers*, interacciona con el dsDNA y la TP (Berman *et al.*, 2007; Kamtekar *et al.*, 2004). El subdominio TPR2 forma, junto con el *palm*, *fingers* y *thumb*, una estructura anular que rodea al dsDNA y estabiliza el complejo en elongación permitiendo la replicación procesiva de todo el genoma a partir de un único evento de asociación con el DNA (Berman *et al.*, 2007; Rodríguez *et al.*, 2005). Por otra parte, el TPR2 forma con el *palm*, *fingers* y dominio exonucleasa un canal estrecho por donde solo puede instalarse un DNA de cadena sencilla. Por su posición relativa, el TPR2 se ha propuesto como la cuña que separa las hebras del DNA y dota a la polimerasa de una capacidad intrínseca de desplazamiento de banda (Berman *et al.*, 2007; Kamtekar *et al.*, 2004; Rodríguez *et al.*, 2005).

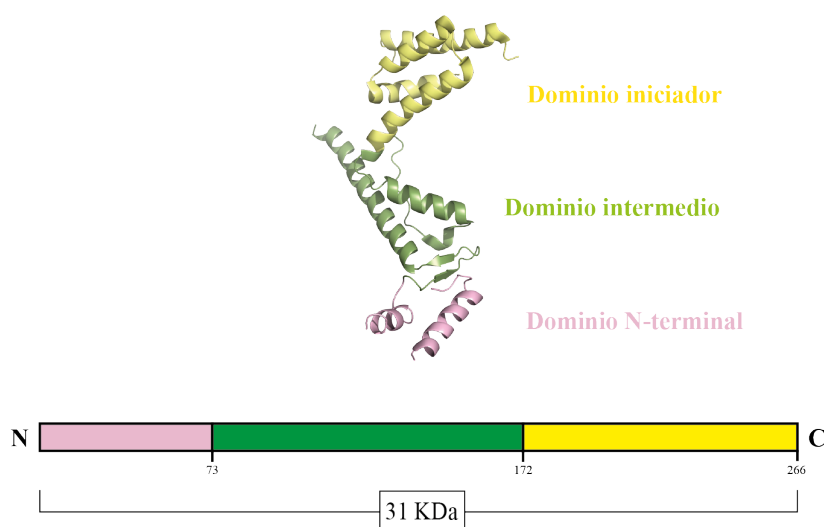
Además de la replicación iniciada con TP (Peñalva y Salas, 1982), la DNA polimerasa de  $\phi 29$  también es capaz de elongar una cadena de DNA a partir de un cebador de DNA (Blanco *et al.*, 1989), lo que la convierte en una potente herramienta para aplicaciones donde no es necesario el mecanismo de replicación cebado por proteína terminal. Recientemente, mediante la inserción de dominios de unión a DNA al extremo C-terminal, se han obtenido DNA polimerasas quiméricas que presentan mayor eficiencia de amplificación (de Vega *et al.*, 2010). Además de su aplicación en los sistemas de amplificación isotérmica como RCA o MDA, la DNA polimerasa de  $\phi 29$  puede ser empleada en el marcaje de extremos con nucleótidos naturales o modificados, la obtención de extremos romos a partir de 3'-recesivos o la generación de sondas de ssDNA a partir de un molde.

### 3.4. La proteína terminal (TP) del bacteriófago $\phi 29$

La proteína terminal es la encargada de proveer el grupo hidroxilo libre necesario para la iniciación de la replicación del genoma de  $\phi 29$  y su mecanismo de acción el responsable de evitar la pérdida de información en los extremos del DNA. Gracias a la resolución de la estructura cristalográfica del heterodímero formado por la DNA polimerasa y la TP, se observó que la TP presentaba una estructura alargada estructurada en tres dominios (Kamtekar *et al.*, 2006): el dominio N-terminal, el dominio intermedio y el dominio C-terminal o iniciador (**Figura 5**).

El dominio N-terminal (residuos 1-73) tiene capacidad de unión a DNA de forma inespecífica de secuencia (Zaballos y Salas, 1989) y dicha unión parece ser la responsable de la localización de la TP en el nucleoide bacteriano, siendo necesario este dominio para una replicación eficiente *in vivo* del DNA viral (Muñoz-Espín *et al.*, 2010) pero parcialmente dispensable para la replicación *in vitro* en presencia de la proteína p6 (Pérez-Arnáiz *et al.*, 2007). Aunque la estructura de este dominio no se pudo resolver a partir del cristal por encontrarse desorganizado, el análisis mediante programas de predicción de estructura se-

cundaria propusieron la existencia de dos hélices- $\alpha$  conectadas por un bucle, viéndose reforzada por los datos obtenidos mediante dicroísmo circular que revelaron un contenido helicoidal aproximadamente del 60% (Holguera *et al.*, 2014). Estudios recientes han demostrado la existencia de una señal de localización nuclear eucariótica (NLS) localizada en los primeros 37 residuos del dominio, además de en fagos como Nf, PRD1, Bam35 y Cp-1, lo cual se propone como un mecanismo de transferencia horizontal de genes entre bacterias y eucariotas (Redrejo-Rodríguez *et al.*, 2012).



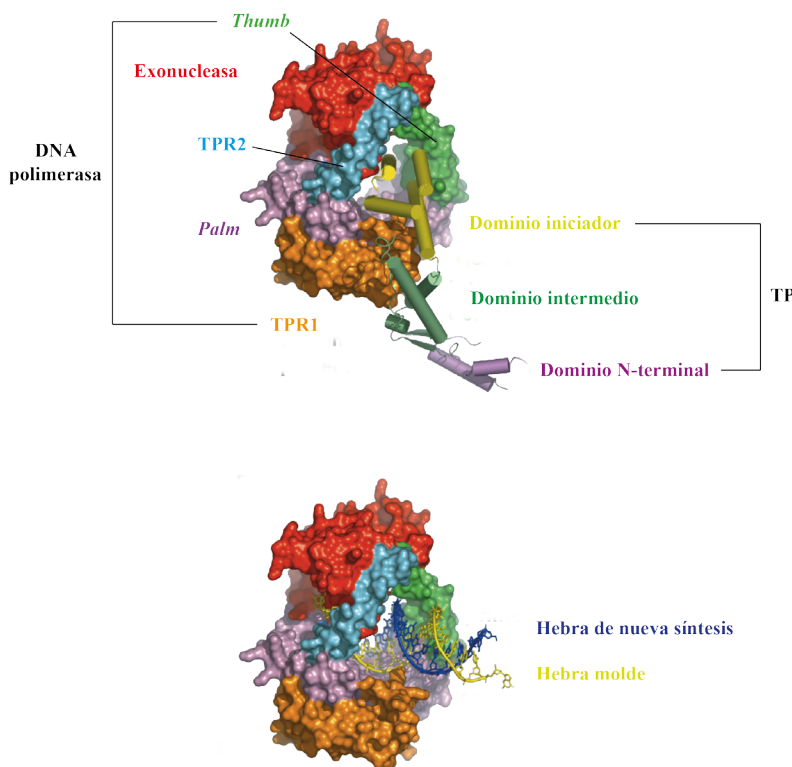
**Figura 5. Estructura tridimensional y mapa de la TP de  $\phi$ 29.** La TP de  $\phi$ 29 es una proteína de 31 KDa con tres dominios diferenciados que se indican de diferente color: un dominio N-terminal (rosa), un dominio intermedio (verde) y un dominio iniciador (amarillo). El modelado se realizó con el software Pymol (<http://www.pymol.org>) a partir de los datos cristalográficos contenidos en el número de acceso 2EX3 de la Protein Data Bank (PDB). Puesto que parte de la estructura del dominio N-terminal no se pudo resolver mediante cristalografía de rayos X, se modeló mediante programas de predicción de estructura secundaria como dos hélices  $\alpha$ .

El dominio intermedio (residuos 74-172) es el responsable del reconocimiento específico entre la DNA polimerasa y la TP mediante su interacción con el subdominio TPR1 de la polimerasa gracias a su carácter electropositivo (Pérez-Arnáiz *et al.*, 2007). Este dominio está conformado por dos hélices- $\alpha$  y se halla conectado al dominio N-terminal por un pequeño giro  $\beta$ , y al dominio C-terminal a través de un bucle flexible que se ha propuesto que actúa como bisagra permitiendo la torsión del dominio iniciador a medida que la DNA polimerasa incorpora los nucleótidos en la hebra naciente. La interacción con el subdominio TPR1 y la presencia del bucle flexible permite una estabilización del complejo durante todo el proceso conocido como transición, hasta que la tensión acumulada tras la incorporación de los primeros nucleótidos



promueve la disociación del heterodímero DNA polimerasa/TP.

El dominio C-terminal o iniciador (residuos 173-266) posee el bucle iniciador que contiene el residuo Ser232, encargado de proveer el grupo 3'OH para la incorporación del primer nucleótido (Hermoso *et al.*, 1985). Aunque este bucle se encuentra desestructurado en el cristal, el dominio consta de cuatro hélices- $\alpha$  y se ha propuesto que su estructura se encarga de mimetizar un dsDNA, mediante su contacto con los subdominios TPR2 y *thumb* de la DNA polimerasa y su carga superficial, ya que ocuparían el mismo lugar en el centro activo de polimerización (Figura 6).



**Figura 6. Estructura tridimensional del heterodímero DNA polimerasa/TP de  $\phi 29$  y posición del DNA molde.** En el panel superior se muestra la estructura tridimensional del heterodímero DNA polimerasa/TP. Se indican los subdominios y dominios de la DNA polimerasa y TP, respectivamente, según los colores mostrados en Kamtekar *et al.*, (2004). En el panel inferior se muestra un modelado de la DNA polimerasa de  $\phi 29$  unida a una estructura tipo DNA molde:iniciador a partir de la estructura terciaria de la DNA polimerasa del bacteriófago RB69. Adaptado de Kamtekar *et al.*, (2006).

### 3.5. La replicación del genoma de $\phi 29$

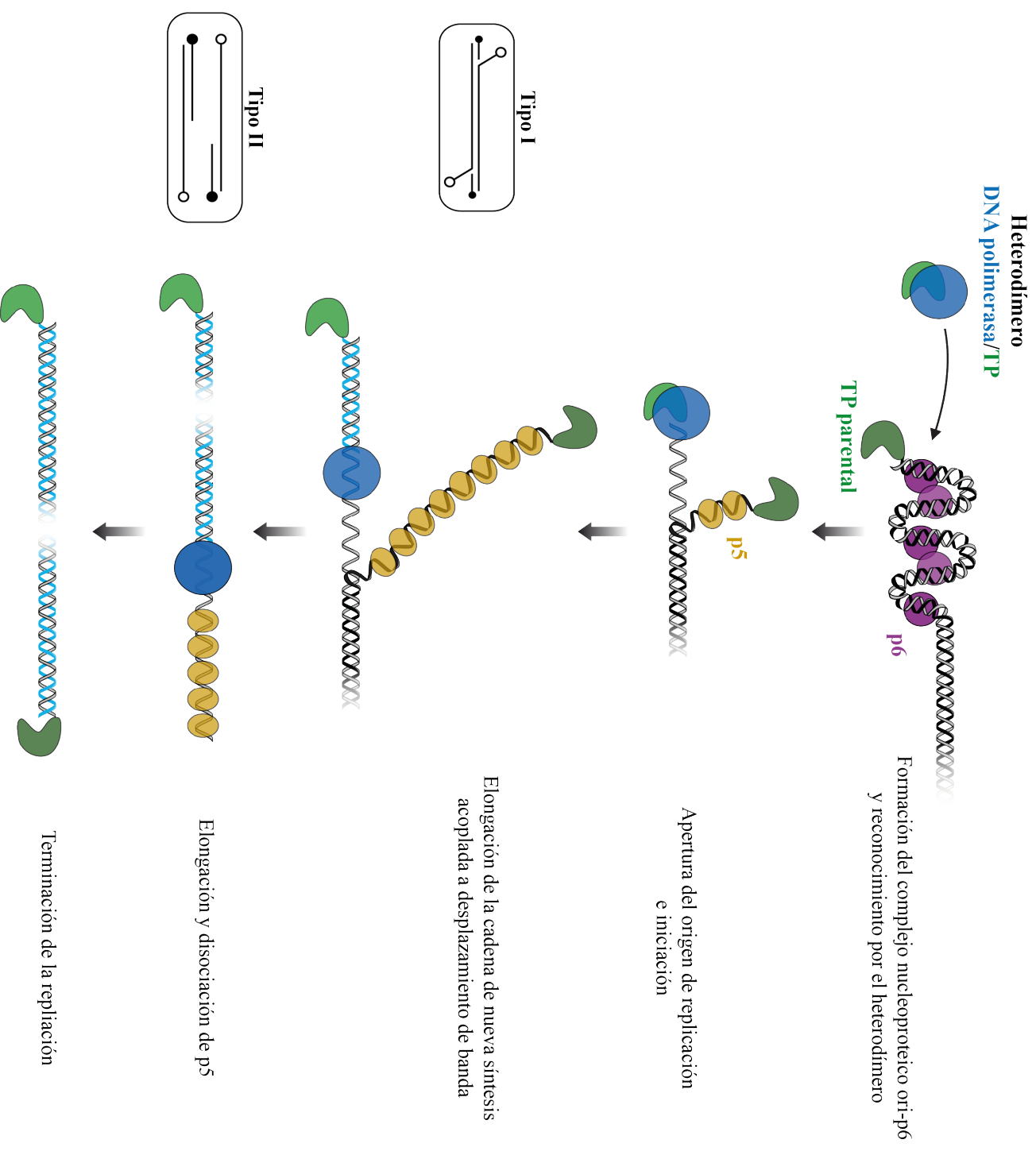
El resultado de la replicación o amplificación del DNA de  $\phi 29$  es la de un DNA lineal con la TP covalentemente unida a sus extremos 5'.

De todos los sistemas replicativos cebados por proteína terminal, el del bacteriófago  $\phi 29$  es el mejor caracterizado gracias a ensayos bioquímicos con proteínas purificadas (Salas y de Vega, 2006) (Figura 7). De acuerdo con el modelo propuesto, la proteína de unión a DNA de doble banda p6 (Prieto *et al.*, 1988)



sería la primera en unirse a los extremos del genoma. Esta unión está mediada por el reconocimiento de un patrón de flexibilidad cada 24 pb en el DNA y genera un complejo nucleoproteico que torsiona el extremo en una superhélice dextrógira (Serrano *et al.*, 1989; revisado en González-Huici *et al.*, 2004b). Esta unión se ha propuesto como un mecanismo de apertura de la hélice de DNA para facilitar la iniciación de la replicación (Prieto *et al.*, 1988; Serrano *et al.*, 1994), así como de reclutamiento de la maquinaria replicativa (complejo DNA polimerasa/TP) (Freire *et al.*, 1996; Pérez-Arnaiz *et al.*, 2007). Como se ha mencionado anteriormente, los orígenes de replicación contienen una TP unida a sus extremos 5' que aumenta la eficiencia de iniciación, paso limitante en la replicación del genoma (Gutiérrez *et al.*, 1986), además de los 12 primeros pb que contienen las ITR (Gutiérrez *et al.*, 1988). Estos orígenes de replicación son reconocidos por el heterodímero formado por una TP y una DNA polimerasa viral libres. El heterodímero se une a ellos y el extremo 3' entra en el bolsillo catalítico para ser utilizado como molde. El inicio de la replicación tiene lugar cuando la DNA polimerasa cataliza la adición del primer nucleótido (dAMP) al grupo OH de la Ser232 de la TP (Blanco y Salas, 1984; Hermoso *et al.*, 1985) utilizando el segundo nucleótido de la cadena molde (Méndez *et al.*, 1992). La adición del dAMP a la TP es  $10^2$ - $10^3$  veces más eficiente que con cualquiera de los otros nucleótidos. De esta forma, la TP mimetiza de forma efectiva la presencia de un cebador en el bolsillo catalítico para el dsDNA y el rol del grupo OH libre es asumido por el de la Ser232 (Kamtekar *et al.*, 2006). Tras la incorporación del primer nucleótido, el complejo TP-dAMP se transloca una posición hacia atrás para que el dAMP incorporado aparee con la primera posición (T), y la segunda pueda volver a ser utilizada como molde. Este mecanismo, que incluye la recolocación del complejo TP-dAMP y evita la pérdida de información en segmentos lineales de DNA, se conoce como *sliding-back* (Méndez *et al.*, 1992). Aunque este proceso fue por primera vez descrito en el sistema replicativo del fago  $\phi 29$ , diversas variantes se han observado en elementos de DNA que contienen TP, como adenovirus (King y van der Vliet, 1994) y los bacteriófagos PRD1 (Caldentey *et al.*, 1992), Nf (González-Huici *et al.*, 1996), GA-1 (Illana *et al.*, 1996) y Cp-1 (Martín *et al.*, 1996). Este tipo de mecanismo requiere secuencias cortas repetidas en los extremos del DNA, como el triplete 5'AAA3' en  $\phi 29$ . En este caso, además de la importancia esencial de la primera y segunda posiciones, la tercera también ha mostrado ser importante en la eficiencia de la iniciación de la replicación (Méndez *et al.*, 1992; González-Huici *et al.*, 2000). Tras la incorporación de los dos primeros nucleótidos, la DNA polimerasa sintetiza un producto de elongación de 5 nucleótidos mientras permanece unida a la TP, gracias al bucle flexible de la TP descrito anteriormente. Una vez que la cadena de nueva síntesis alcanza una longitud de 6-9 nucleótidos durante la fase conocida como 'transición', el heterodímero se disocia (Méndez *et al.*, 1997) y la DNA polimerasa replica de forma procesiva y acoplada al desplazamiento de banda las 19 Kb del genoma viral en un proceso idéntico al de una DNA polimerasa dependiente de DNA. Ambos extremos del genoma son empleados como orígenes de replicación, de forma que la duplicación del genoma se da en un proceso de replicación simétrica

**Figura 7. Esquema del proceso de replicación *in vitro* del genoma de  $\phi 29$ .** De acuerdo con el modelo propuesto, la proteína p6 se une a ambos extremos del genoma formando un complejo nucleoproteico que, mediante la torsión de la doble cadena, facilitaría la apertura de la hélice. Esta apertura permitiría el acceso y reconocimiento de los orígenes de replicación por parte del heterodímero formado por la DNA polimerasa y la TP. A continuación, la DNA polimerasa cataliza la adición del primer dAMP al grupo OH del residuo S232 de la TP. Tras la etapa de transición (no representado), el heterodímero se disocia, permitiendo a la DNA polimerasa continuar con la elongación de la cadena de nueva síntesis de forma procesiva y acoplada a desplazamiento de banda. Esto genera los intermediarios replicativos de tipo I y II. La proteína p5 va a ir uniéndose a las hebras de ssDNA desplazadas evitando así su degradación por parte de las nucleasas. El proceso de replicación concluye con la disociación de la DNA polimerasa del DNA tras la incorporación del último nucleótido.



pero no necesariamente simultánea (Blanco *et al.*, 1989). Esto, unido al desplazamiento de banda, hace que durante el proceso se generen intermediarios replicativos, compuestos por una molécula de dsDNA con una o dos colas de ssDNA, de tipo I, mientras las hebras molde permanecen parcialmente hibridadas (Harding e Ito, 1980; Inciarte *et al.*, 1980; Sogo *et al.*, 1982), y de tipo II (Inciarte *et al.*, 1980; Sogo *et al.*, 1982), cuando ambas hebras molde están completamente desapareadas. Durante el proceso, la proteína viral p5 (SSB) se une a la cadena sencilla de la hebra desplazada, protegiéndola de la acción de nucleasas (Martin *et al.*, 1989) y evitando la formación de estructuras secundarias y uniones no productivas de la polimerasa (Soengas *et al.*, 1995), además de estimular la replicación (Gutiérrez *et al.*, 1991). Conforme la polimerización progresa, estas SSB se van disociando con el avance de la polimerasa. Una vez que la DNA polimerasa alcanza el extremo opuesto del genoma, tiene lugar la terminación de la replicación, con su disociación del DNA y la generación de las dos copias de DNA.

A pesar de que el mecanismo de replicación cebado por TP es un mecanismo común en la familia de fagos relacionados con  $\phi 29$ , hay ciertas diferencias en el proceso de recuperación de la información genética mediada por TP. Al igual que en  $\phi 29$ , la reacción de iniciación en el caso del fago GA-1 está dirigido por el segundo nucleótido de la hebra molde (Illana *et al.*, 1996). Sin embargo, en Nf o Cp-1 está dirigido por el tercero (Martín *et al.*, 1996) y en PRD1 por el cuarto (Caldentey *et al.*, 1993), recuperando la información mediante sucesivos pasos de *sliding-back* en un proceso conocido como *stepwise sliding-back* (Martín *et al.*, 1996). Por otra parte, en el caso de adenovirus, cuya secuencia en el extremo es 3'GTAG-TA<sup>5'</sup>, el heterodímero sintetiza el triplete CAT tomando como molde las posiciones 4-6 y posteriormente, en un mecanismo conocido como *jumping-back*, este apareará con las posiciones 1-3 y procederá con la elongación de la cadena (King y van der Vliet, 1994). Estudios recientes en el sistema de  $\phi 29$  indican que el residuo Phe230 de la TP cebadora parece ser uno de los principales implicados en marcar la posición del nucleótido que va a ser tomado como molde para la iniciación (del Prado, *et al.*, 2015). Por otra parte, estudios comparativos realizados con los fagos Nf y GA-1 mostraron que, a pesar de la homología entre sus proteínas, hay una gran especificidad en las interacciones de la DNA polimerasa, la TP y los orígenes de replicación (González-Huici *et al.*, 2000), lo que explicaría la existencia de las diferencias observadas en el mecanismo de iniciación.

### 3.6. Una herramienta biotecnológica

El sistema replicativo del bacteriófago  $\phi 29$  ha demostrado ser un ejemplo en la aplicación de la investigación básica en el sector biotecnológico. La base biotecnológica del sistema se fundamenta en su DNA polimerasa y sus propiedades de alta procesividad y capacidad de desplazamiento de banda, que la convierten en el enzima idóneo para aplicaciones como la amplificación de ácidos nucleicos. Además, la

inserción de dominios de unión a DNA ha permitido obtener DNA polimerasas quiméricas con una mayor eficiencia y productividad que el enzima *wild-type*, con las que se conseguía hasta 2.7 veces más producto con la técnica de MDA (de Vega *et al.*, 2010). Su repercusión puede verse reflejada en la comercialización de kits de amplificación de ácidos nucleicos basados en las técnicas de RCA (Dean *et al.*, 2001) o MDA (Dean *et al.*, 2002), así como en su aplicación en la secuenciación masiva de genomas (WGA) y metagenómica. Por otra parte, sus características también son de gran utilidad para las aplicaciones derivadas de la nueva Biología Sintética como en el campo de los DNAs no naturales (XNA, Ácido Xenonucleico) (Pinheiro *et al.*, 2012).

El sistema mínimo de amplificación *in vitro* fue desarrollado por el laboratorio en base a 4 proteínas purificadas que son codificadas por el propio fago: la DNA polimerasa, la TP, la proteína p5 y la proteína p6. Este es un sistema altamente eficiente que permite obtener una amplificación del TP-DNA de 19 Kb de hasta 1000 veces manteniendo su funcionalidad (Blanco *et al.*, 1994). El resultado de la amplificación es un DNA unido covalentemente a una TP en cada extremo. Esto hace del sistema un punto de partida muy interesante para tecnologías como el DNA display, facilitando sustancialmente la unión genotipo-fenotipo ya que, en el proceso de replicación, una proteína (TP) queda unida covalentemente al DNA que contiene su gen codificante (revisado en Sergeeva *et al.*, 2006). Esta unión genotipo-fenotipo también puede emplearse como base en la técnica de compartimentalización *in vitro* (IVC), basada en la encapsulación de genes y una maquinaria de transcripción-traducción que permite la selección de propiedades determinadas en las proteínas codificadas por su respectivo gen (Tawfik y Griffiths, 1998). En este sentido, el hecho de que el DNA de  $\phi 29$  codifique para las proteínas replicativas permitiría el desarrollo de un sistema de compartimentalización autorreplicante.

La unión covalente de la TP también podría servir como plataforma para fusionar una amplia variedad de módulos proteicos que tuvieran propiedades catalíticas, de unión, interacción o señalización. De hecho, gracias al reciente hallazgo de una señal de localización nuclear eucariota (NLS) en el dominio N-terminal de la TP de  $\phi 29$  y otros bacteriófagos (Redrejo-Rodríguez *et al.*, 2012; Redrejo-Rodríguez y Salas, 2014), se ha propuesto el sistema de  $\phi 29$  como herramienta potencial de translocación nuclear de genes en células de mamífero en el contexto más actual de la terapia génica. Además, la posibilidad de añadir funciones al DNA también posibilitaría su integración con la tecnología de aptámeros de DNA, que permitiría fusionar una actividad determinada a un DNA con capacidad de reconocimiento de diana.

Por otra parte, los estudios básicos sobre los procesos de replicación, transcripción y de interacción virus-hospedador han permitido que los conocimientos generados a partir del estudio del sistema de  $\phi 29$  puedan ser extrapolables a otros sistemas virales de interés económico y sanitario. En algunos casos, porque estos virus comparten una organización genómica o mecanismo replicativo similar a  $\phi 29$ , como algunos virus de plantas o el adenovirus humano (revisado en Salas, 1991), que puede producir transfor-

mación oncogénica (Endter *et al.*, 2004). En otros casos, porque la comprensión de los mecanismos del ciclo viral permiten el potencial desarrollo de plataformas terapéuticas, como el reconocimiento por parte del hospedador (Holguera *et al.*, 2012) en el desarrollo de inhibidores del ciclo infectivo de otros virus, o las propiedades del pRNA de  $\phi 29$ , implicado en el empaquetamiento viral, como base para ribozimas (Hoeprich *et al.*, 2003).



# Objetivos





El sistema replicativo del bacteriófago  $\phi 29$  ha tenido un gran impacto en el sector de la biología aplicada como herramienta biotecnológica. Sin embargo, la aplicabilidad del sistema no puede ser desligada del conocimiento básico de sus procesos. Por esa razón, y con el objetivo de proveer de nuevas aplicaciones al sistema replicativo de  $\phi 29$  y su maquinaria de replicación, se han aunado ambos aspectos en la consecución de los siguientes objetivos:

1. Desarrollar un sistema de amplificación cebado por TP de DNAs heterólogos que contienen los orígenes de replicación del DNA del bacteriófago  $\phi 29$  basado en el sistema de replicación del propio fago.
2. Mejorar la eficiencia de los orígenes de replicación del DNA de  $\phi 29$  y realizar un estudio sistemático y comprensivo de las primeras fases del proceso replicativo del bacteriófago  $\phi 29$ .
3. Mejorar la eficiencia de utilización de los orígenes de replicación del bacteriófago  $\phi 29$  mediante el reclutamiento específico de la maquinaria de replicación al origen: sustitución del dominio N-terminal de la TP de  $\phi 29$  por el dominio de unión a DNA de GAL4.
4. Identificar sitios potenciales de inserción de secuencias peptídicas en la TP del bacteriófago  $\phi 29$  para la incorporación de nuevas funciones.



# Materiales y Métodos



## 1. Materiales

### 1.1. Nucleótidos y oligonucleótidos

#### 1.1.1. Nucleótidos

Los dNTPs no marcados radiactivamente y los 2',3'-ddNTPs se obtuvieron de GE Healthcare (pureza trifosfato de >99% y  $\geq 98\%$ , respectivamente). El [ $\gamma$ - $^{32}\text{P}$ ] ATP (3000 Ci/mmol) y los [ $\alpha$ - $^{32}\text{P}$ ] dNTPs (3000 Ci/mmol) se obtuvieron de Perkin Elmer.

#### 1.1.2. Oligonucleótidos

Los oligonucleótidos empleados en la presente tesis se obtuvieron de Sigma-Aldrich. La secuencia de los oligonucleótidos que no se emplearon para la construcción de los orígenes de replicación artificiales se detalla en la **Tabla 1**. Los orígenes artificiales dsDNA de 29 y 68 pb del genoma de  $\phi 29$  se obtuvieron mediante la hibridación de los oligonucleótidos correspondientes, previamente purificados en el laboratorio (véase apartado 2.2). En este caso, todas las construcciones llevan un grupo fosfato en el extremo 5' de la hebra desplazada. La secuencia de estos orígenes se detalla en la **Tabla 2**.

**Tabla 1. Oligonucleótidos**

Nombre	Secuencia (5'→3')
<b>BaeIFw</b>	[Phos]TCTCCTAATTATTTGACTGTCGTATCTGTCATCAACCA
<b>BaeIRev</b>	[Phos]TGGTTGATGACAGATACGACAGTCAAATAATTAGGAGA
<b>BsmBIFw</b>	GTTATCCCCTGCGTCTCTGGATAACCGTAT
<b>BsmBIRev</b>	ATACGGTTATCCAGAGACGCAGGGGATAAC
<b>68L</b>	[Phos]AAAGTAAGCCCCACCCTCACATGATACCATTCTCCTAATATCGACATAATCCGTCGATCCTCGGCAT
<b>68LEco</b>	[Phos]AATTATGCCGAGGATCGACGGATTATGTCGATATTAGGAGAATGGTATCATGTGAGGGTGGGGGCTTACTTT
<b>p2Eco</b>	CTTCTCTTGATCGAAGAATTCACCCATTAC
<b>p3Nco</b>	GGAGATAACGCACCATGGCGAGAAGTCC
<b>p6Nco</b>	GGGACGAAGCCATGGCAAAAATGATG
<b>pETORPHIBsmRev</b>	TAGCAGCCGCGTCTCAGGATTGGTGGTGGT
<b>T7Term</b>	TAATACGACTCACTATAGGG
<b>T7Univ</b>	TAATACGACTCACTATAGGG
<b>Tn5Fw</b>	CGTCAAGGCCGATGGTACCA
<b>Tn5Rev</b>	CCCATGAGGCCAGGAGCTCAG
<b>GFPup</b>	CTTTCATAACCGTGCTAATATGC
<b>GFPdown</b>	GAAGTGGATCCAGTTAGTATAGTTCATCC

Tabla 2. Orígenes artificiales basados en los extremos del genoma de  $\phi 29$

Nombre	Secuencia
15	3' TTTTCATCCCATGTCG <sup>5'</sup>
17	3' TTTTCATCCCATGTCGCT <sup>5'</sup>
19	3' TTTTCATCCCATGTCGCTGT <sup>5'</sup>
21	3' TTTTCATCCCATGTCGCTGTTG <sup>5'</sup>
23	3' TTTTCATCCCATGTCGCTGTTGTA <sup>5'</sup>
25	3' TTTTCATCCCATGTCGCTGTTGTATG <sup>5'</sup>
27	3' TTTTCATCCCATGTCGCTGTTGTATGTG <sup>5'</sup>
A/T (wt/wt <sub>R</sub> /wt <sub>ds</sub> )	5' AAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTTCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
G/T	5' GGGGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTTCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
T/T	5' TTTGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTTCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
C/T	5' CCCGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTTCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
A/C	5' AAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
G/C	5' GGGGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
T/C	5' TTTGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
C/C	5' CCCGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
A/A	5' AAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' AAACATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
G/A	5' GGGGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' AAACATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
T/A	5' TTTGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' AAACATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
C/A	5' CCCGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' AAACATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
A/G	5' AAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' GGGCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
G/G	5' GGGGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' GGGCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
T/G	5' TTTGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' GGGCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
C/G	5' CCCGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' GGGCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
wt <sub>L</sub>	5' AAAGTAAGCCCCACCCTCACATGATACC <sup>3'</sup> 3' TTTTCATTCGGGGTGGGAGTGTACTATGG <sup>5'</sup>
GA-1 <sub>L</sub>	5' AAATAGAGTCCACACCCTCACATGATACC <sup>3'</sup> 3' TTTATCTCAGGTGTGGGAGTGTACTATGG <sup>3'</sup>
A/T <sup>TAG</sup>	5' AAATAGGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTATCCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
A/T <sup>TAG</sup> <sub>CAT</sub>	5' AAATAGGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTTCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>

<b>A/T<sup>GTA</sup><sub>ATC</sub></b>	5' AAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTATCCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>A/T<sup>TCA</sup></b>	5' AAATCAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTAGTCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>A/T<sup>AGT</sup></b>	5' AAAAGTGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTTCACCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>A/T<sup>CGA</sup></b>	5' AAACGAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTGCTCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>A/T<sup>CAT</sup></b>	5' AAACATGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTGTACCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>G/C<sup>TAG</sup><sub>CAT</sub></b>	5' GGGTAGGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>A/C<sup>TAG</sup><sub>CAT</sub> (6m)</b>	5' AAATAGGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>wt<sub>ss</sub></b>	3' TTTTCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>A/ddC</b>	5' AAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' ddCCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>G/ddC</b>	5' GGGTAGGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' ddCCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>10+A/ddC</b>	5' ATCACTATGCAAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' ddCCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>10+G/ddC</b>	5' ATCACTATGCGGGGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' ddCCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>10+TAC/ddC</b>	5' ATCACTATGCTACGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' ddCCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>AAA10</b>	5' AAAACTATGCTACGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' ddCCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>5'hA/T</b>	5' CGCTATCTTAGCGAAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTTCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>5'Δ6wt</b>	5' GGGTACAGCGACAACATACACCA <sup>3'</sup> 3' TTTTCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>AAC</b>	5' AACGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>ACA</b>	5' ACAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>CAA</b>	5' CAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>ACC</b>	5' ACCGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>CAC</b>	5' CACGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>CCA</b>	5' CCAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>CCC</b>	5' CCGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>8m</b>	5' AAATAGACGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>10m</b>	5' AAATAGACTCACAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>12m</b>	5' AAATAGACTCGTAGCGACAACATACACCA <sup>3'</sup> 3' CCCCATCCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>AAA 6m'</b>	5' AAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCGTACCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>
<b>CAA 6m'</b>	5' CAAGTAGGGTACAGCGACAACATACACCA <sup>3'</sup> 3' CCGTACCCATGTCGCTGTTGTATGTGGT <sup>5'</sup>

<b>wt (68)</b>	5' AAAGTAGGGTACAGCGACAACATACACCATTTCCTCCATTGACCGACTATCTTCGACAAGAATCTAACA3' 3' TTTTCATCCCATGTCGCTGTTGTATGTGGTAAAGGGGTAAGTGGCTGATAGAAGCTGTTCTTAGATTGT5'
<b>G/C (68)</b>	5' GGGGTAGGGTACAGCGACAACATACACCATTTCCTCCATTGACCGACTATCTTCGACAAGAATCTAACA3' 3' TTTTCATCCCATGTCGCTGTTGTATGTGGTAAAGGGGTAAGTGGCTGATAGAAGCTGTTCTTAGATTGT5'
<b>A/C (68)</b>	5' AAAGTAGGGTACAGCGACAACATACACCATTTCCTCCATTGACCGACTATCTTCGACAAGAATCTAACA3' 3' CCCCATCCCATGTCGCTGTTGTATGTGGTAAAGGGGTAAGTGGCTGATAGAAGCTGTTCTTAGATTGT5'
<b>6m (68)</b>	5' AAATAGGGGTACAGCGACAACATACACCATTTCCTCCATTGACCGACTATCTTCGACAAGAATCTAACA3' 3' CCCCATCCCATGTCGCTGTTGTATGTGGTAAAGGGGTAAGTGGCTGATAGAAGCTGTTCTTAGATTGT5'
<b>8m (68)</b>	5' AAATAGACGTACAGCGACAACATACACCATTTCCTCCATTGACCGACTATCTTCGACAAGAATCTAACA3' 3' CCCCATCCCATGTCGCTGTTGTATGTGGTAAAGGGGTAAGTGGCTGATAGAAGCTGTTCTTAGATTGT5'
<b>10m (68)</b>	5' AAATAGACTCACAGCGACAACATACACCATTTCCTCCATTGACCGACTATCTTCGACAAGAATCTAACA3' 3' CCCCATCCCATGTCGCTGTTGTATGTGGTAAAGGGGTAAGTGGCTGATAGAAGCTGTTCTTAGATTGT5'
<b>12m (68)</b>	5' AAATAGACTCGTAGCGACAACATACACCATTTCCTCCATTGACCGACTATCTTCGACAAGAATCTAACA3' 3' CCCCATCCCATGTCGCTGTTGTATGTGGTAAAGGGGTAAGTGGCTGATAGAAGCTGTTCTTAGATTGT5'
<b>UAS</b>	5' AAAGTAGGGTACAGCGGCAACATACACCCGGTCCCTCCATTGACCGACTATCTTCGACAAGAATCTAACA3' 3' TTTTCATCCCATGTCGCTGTTGTATGTGGTAAAGGGGTAAGTGGCTGATAGAAGCTGTTCTTAGATTGT5'
<b>68LEcoO109I</b>	5' AAAGTAAGCCCCACCCTCACATGATACCATTCTCCTAATATCGACATAATCCGTGATCCTCGGCAT3' 3' TTTTCATTCGGGGTGGGAGTGTACTATGGTAAGAGGATTATAGCTGTATTAGGCAGCTAGGAGCCGTACCG5'
<b>68LBsmBI</b>	5' AAAGTAAGCCCCACCCTCACATGATACCATTCTCCTAATATCGACATAATCCGTGATCCTCGGCAT3' 3' TTTTCATTCGGGGTGGGAGTGTACTATGGTAAGAGGATTATAGCTGTATTAGGCAGCTAGGAGCCGTACCTA5'



## 1.2. Plásmidos y construcciones lineales

Los plásmidos y construcciones lineales empleados en la presente tesis junto con la forma en que se han obtenido, tamaño y uso se detallan en la **Tabla 3**.

**Tabla 3. Plásmidos y construcciones lineales**

Nombre	Construcción o procedencia	Tamaño (pb)	Uso
<b>TP-DNA del bacteriófago <math>\phi</math>29</b>	Obtenido como se describe en Peñalva y Salas (1982)	19285	Análisis de actividad de replicación
<b>Plásmido pT7-3</b>	Perteneciente a la serie pT-7 (Tabor y Richardson, 1985)	2423	Plásmido general de expresión
<b>Plásmido pT7-TP2ST</b>	Obtenido mediante el clonaje del gen 3 del genoma de $\phi$ 29, correspondiente a la TP, y un doble Strep-tag en su extremo N-terminal, entre las dianas para los enzimas de restricción EcoRI y BamHI del plásmido pT7-3. El gen queda bajo el control del promotor $\phi$ 10 del bacteriófago T7	3066	Plásmido empleado para la expresión de la TP Strep (que contiene un doble Strep-tag) y plataforma para la construcción de la librería de inserciones de GFP en la TP de $\phi$ 29
<b>Plásmido pET28b</b>	Obtenido de Novagen	5368	Plásmido general de clonaje
<b>Plásmido pEYFP-N1</b>	Obtenido de Clontech	4700	Plásmido empleado para la expresión de la proteína YFP en células eucariotas
<b>Plásmido pGFP-NlaKmTrans</b>	Cedido amablemente por Belén Calles y Víctor de Lorenzo (Billerbeck <i>et al.</i> , 2013)	4285	Plásmido que contiene un derivado del transposón Tn5 para la transposición al azar de GFP en el gen de la TP de $\phi$ 29
<b>Plásmido pUC57ORPHI</b>	Obtenido de Genescript mediante el clonaje en un plásmido pUC57 de un fragmento de DNA correspondiente a los primeros 191 y 194 pb de los extremos izquierdo y derecho del genoma de $\phi$ 29, respectivamente, unidos de forma invertida. La unión de los extremos genera una diana de restricción para el enzima DraI	3095	Plásmido general de clonaje
<b>Plásmido pETORPHI</b>	Obtenido mediante el clonaje entre las dianas para los enzimas de restricción SacI y HindIII de un plásmido pET28b del fragmento de DNA generado tras la digestión con las mismas enzimas del plásmido pUC57ORPHI. La digestión del plásmido pETORPHI permite obtener un DNA lineal flanqueado por las secuencias de los extremos del genoma de $\phi$ 29	5833	Plásmido empleado para el desarrollo del sistema de amplificación de DNAs heterólogos cebado por TP y análisis de diferentes actividades

<b>Plásmido pETORPHI68L</b>	Obtenido mediante la digestión del plásmido pETORPHI con los enzimas de restricción DraI y EcoRI. En su lugar, se clonó un fragmento de 68 pb (hibridación de los oligonucleótidos 68L y 68LEco), correspondiente a la secuencia del extremo izquierdo del DNA de $\phi 29$	5707	Plásmido basado en el pETORPHI empleado para analizar las secuencias mínimas de estimulación por p6 en el sistema de amplificación de DNAs heterólogos cebado por TP
<b>Plásmido pETORPHIBae</b>	Obtenido mediante el clonaje en la posición de la diana para el enzima de restricción DraI de un fragmento de dsDNA (hibridación de los oligonucleótidos BaeIFw y BaeIRev) que contenía la secuencia del sitio de reconocimiento para el enzima de restricción BaeI. Los sitios de corte para BaeI se sitúan inmediatamente adyacentes a los extremos 5' de los extremos del DNA de $\phi 29$	5302	Plásmido basado en el pETORPHI para el análisis del efecto de la linealización con el enzima de restricción BaeI, menos frecuente que DraI y con extremos 3' protuberantes
<b>Plásmido pEYFPBsm</b>	Obtenido mediante la amplificación del plásmido pEYFP-N1 con los oligonucleótidos BsmBIFw y BsmBIRev, generando una diana para el enzima de restricción BsmBI en la posición 4702	4702	Plásmido basado en el pEYFP-N1
<b>Plásmido pEYFPORBae</b>	Obtenido mediante el clonaje entre las dianas para los enzimas de restricción NdeI y BsmBI del plásmido pEYFPBsm de los orígenes de $\phi 29$ del plásmido pETORPHI. Los orígenes se extrajeron mediante la amplificación del plásmido pETORPHIBae con los oligonucleótidos T7Univ y pETORPHIBsmRev	5302	Plásmido basado en el pEYFP-N1 para la amplificación cebada por TP con un plásmido que expresa la proteína YFP en células eucariotas

### 1.3. Proteínas y péptidos

La **DNA polimerasa *wild-type*** del bacteriófago  $\phi 29$  y el mutante deficiente en actividad exonucleasa **D12A/D66A** (Bernad et al., 1989) se purificaron a partir de su expresión en células de *Escherichia coli* como se describe en Lázaro et al. (1995).

La **TP *wild-type*** del bacteriófago  $\phi 29$  y la **TP GAL4** se purificaron como se describe en Mencía et al. (2011). El mutante de la TP en el residuo iniciador **S232C** y el de delección del dominio N-terminal **TP  $\Delta N$**  se purificaron como se describe en Garmendia et al. (1990).

Las TP que contenían un Strep-tag en su extremo N-terminal: la **TP Strep**, las TPs con insertos (**i38**, **i65**, **i70**, **i75**, **i115**, **i118**, **i144**, **i174**, **GFP14** y **GFP38**), la **TP GFP** y la **YFP TP** se purificaron con el kit de purificación Strep tag-Strep tactin de IBA siguiendo las instrucciones del fabricante a partir de colonias de *E.coli* cultivadas en 200 ml y condiciones de autoinducción a 30°C durante la noche.

La **p5** del bacteriófago  $\phi 29$  se obtuvo a partir de células de *B.subtilis* infectadas con  $\phi 29$  y se purificó como se describe en Soengas et al. (1995).

La **p6** del bacteriófago  $\phi 29$  se obtuvo a partir de la expresión en células de *E.coli* de un plásmido pT-7 (Tabor y Richardson, 1985) que contenía el gen de la p6 y se purificó como se describe en Pastrana et al. (1985).

Las **endonucleasas de restricción** y **enzimas de modificación** se obtuvieron de *New England Biolabs*.

La **transposasa Tn5** se purificó como se describe en Bhasin et al., (1999).

Los péptidos de penetración celular **TAT** (AGYGRKKRRQRRR) y **ANTP** (SGRQIKIWFQNRR-MKWKK) unidos a ácido 3-maleimidopropiónico (Mal-TAT y Mal-ANTP, respectivamente) se obtuvieron de GenScript a una pureza de  $\geq 95\%$  y se disolvieron en agua ultrapura.

### 1.4. Estirpes bacterianas y líneas celulares de mamífero

***E.coli* XL-1 Blue**: Las bacterias *E.coli* XL-1 Blue se emplearon para los procesos de clonaje y se obtuvieron de Agilent.

***E.coli* BL21(DE3)**: Las bacterias *E.coli* BL21(DE3), que expresa la RNA polimerasa de T7, se emplearon para los procesos de expresión de proteínas y posterior purificación y se obtuvieron de Novagen.

### 1.5. Medios de cultivo

Las bacterias *E.coli* se cultivaron en medio de cultivo Luria Bertani (LB) con el antibiótico requerido en cada caso: ampicilina (100  $\mu\text{g/ml}$ ) o kanamicina (100  $\mu\text{g/ml}$ ).

## 2. Métodos

### 2.1. Ensayo de amplificación en presencia de dATP marcado radiactivamente

La mezcla de reacción contenía, en 25  $\mu$ l, 50 mM Tris-HCl, pH 7.5, 10 mM  $MgCl_2$ , 4% (v/v) glicerol, 1 mM DTT, 0.1 mg/ml BSA, 20 mM  $(NH_4)_2SO_4$ , 100  $\mu$ M dNTPs y 1  $\mu$ Ci (13.4 nM)  $[\alpha\text{-}^{32}P]$ dATP (ratio  $[\alpha\text{-}^{32}P]$ dATP:dNTPs = 1:7400). Las concentraciones de reacción de las 4 proteínas purificadas del sistema replicativo de  $\phi$ 29, salvo cuando se especifica lo contrario, fueron: 6 nM de DNA polimerasa, 25 nM de TP, 23  $\mu$ M de p5 y 5  $\mu$ M, en el caso del plásmido pETORPHI, o 30  $\mu$ M, en el caso del TP-DNA de  $\phi$ 29, de p6, así como las cantidades indicadas en cada experimento de molde de DNA. La DNA polimerasa y la TP de  $\phi$ 29 fueron incubadas juntas en hielo durante 15 minutos antes de la reacción. Las reacciones se incubaron durante 2 horas a 22°C y se detuvieron mediante la adición de 10 mM EDTA y 0.1% (p/v) SDS, concentraciones finales. Las muestras se filtraron en columnas de Sephadex G-50 equilibradas en 10 mM Tris-HCl, pH 7.5, 1 mM EDTA y 0.1% SDS en una centrífuga Thermo Scientific Heraeus Labofuge 400 a 2800 rpm durante 5 minutos con el objeto de eliminar el  $[\alpha\text{-}^{32}P]$ dATP no incorporado. Posteriormente, se secaron para ser resuspendidas en 50  $\mu$ l de una solución de NaOH 0.5 M y se analizaron mediante electroforesis en geles de agarosa al 0.7% (p/v) en presencia de 30 mM NaOH y 1 mM EDTA en el mismo gel y tampón de electroforesis. Los geles se secaron al vacío a 80°C. El revelado se realizó por autorradiografía. La cuantificación de las bandas del gel se realizó por medio de un equipo Fujifilm BAS1000 phosphorimager y el DNA sintetizado se determinó mediante el cálculo del ratio entre la radiactividad incorporada con respecto al total de la radiactividad presente inicialmente en la reacción. Este ratio se multiplicó por la masa inicial de dNTPs presente inicialmente en la reacción para el cálculo de la masa total de dNMPs incorporados al DNA. Como marcador de peso molecular se utilizó el DNA del fago lambda digerido con HindIII y marcado radiactivamente mediante su incubación con el enzima Klenow y 50  $\mu$ M de cada dCTP, dGTP, dTTP y 1  $\mu$ Ci (13.4 nM)  $[\alpha\text{-}^{32}P]$ dATP en el tampón EcoPol (*New England Biolabs*) durante 25 minutos a 25°C.

### 2.2. Purificación de oligonucleótidos

Los oligonucleótidos para la construcción de los orígenes artificiales de  $\phi$ 29 fueron purificados en el laboratorio a partir de geles de poliacrilamida al 20% y 8 M urea como se describe en Maniatis *et al.*, (1982).

### 2.3. Hibridación de oligonucleótidos

Los oligonucleótidos para la construcción de los orígenes artificiales de  $\phi$ 29 se hibridaron en presencia de 60 mM Tris-HCl, pH 7.5 y 200 mM NaCl. La desnaturalización y rehibridación se realizó en un termociclador con el siguiente programa: 95°C, 5 minutos; 93°C, 30 segundos; 85°C, 10 minutos; 80°C, 10 minutos; 75°C, 5 minutos; 70°C, 5 minutos; 65°C, 5 minutos; 45°C 60 minutos; 4°C.

#### 2.4. Ensayo de iniciación de la replicación (formación del complejo TP-dAMP)

La mezcla de reacción contenía, en 25  $\mu$ l, 50 mM Tris-HCl, pH 7.5, 10 mM  $MgCl_2$ , 4% (v/v) glicerol, 1 mM DTT, 0.1 mg/ml BSA, 20 mM  $(NH_4)_2SO_4$  y 0.1  $\mu$ M y 1  $\mu$ Ci (13.4 nM) del correspondiente dNTP y  $[\alpha\text{-}^{32}P]$ dNTP, respectivamente (ratio  $[\alpha\text{-}^{32}P]$ dATP:dATP = 1:7.5). Las concentraciones de las proteínas purificadas del sistema replicativo de  $\phi$ 29 en la reacción, salvo cuando se especifica lo contrario, fueron: 30 nM de DNA polimerasa, 150 nM TP, así como 0.63 nM  $\phi$ 29 TP-DNA o 313 nM del oligonucleótido de doble banda especificado. La DNA polimerasa y la TP de  $\phi$ 29 fueron incubadas juntas en hielo durante 15 minutos antes de la reacción. Para el análisis del efecto de p6 en oligonucleótidos dsDNA de 68 pb, se añadieron 10.5  $\mu$ g de p6 donde se indica. Las reacciones se incubaron durante 5 minutos a 30°C y se detuvieron mediante la adición de 10 mM EDTA y 0.1% (p/v) SDS, concentraciones finales. Las muestras se filtraron como se indica en el *Apartado 2.1*. Posteriormente, se secaron y fueron resuspendidas en tampón de carga de proteínas (37 mM Tris-HCl pH 6.8, 2% (p/v) SDS, 4% (v/v) 2-mercaptoetanol, y 13% (v/v) glicerol) para ser analizadas por SDS-PAGE en geles de poliácridamida al 12% (Peñalva y Salas, 1982). El revelado se realizó por autorradiografía. La cuantificación de las bandas del gel se realizó por medio de un equipo Fujifilm BAS1000 phosphorimager. Las medias y desviaciones estándar que se muestran son el resultado de la cuantificación de, al menos, tres experimentos independientes. La cantidad de DNA producido por cada construcción de doble cadena se normalizó considerando la cantidad de DNA producido por el origen wt como 1. Como marcador de la posición +1 en los experimentos con oligonucleótidos de doble banda, se realizó un experimento de iniciación de la replicación con TP-DNA como molde, además de como forma de estandarización adicional de los experimentos independientes.

#### 2.5. Ensayo de transición con oligonucleótidos dsDNA

La mezcla de reacción contenía, en 25  $\mu$ l, 50 mM Tris-HCl, pH 7.5, 10 mM  $MgCl_2$ , 4% (v/v) glicerol, 1 mM DTT, 0.1 mg/ml BSA, 20 mM  $(NH_4)_2SO_4$  y 313 nM del oligonucleótido de doble banda especificado. Las concentraciones de reacción de las proteínas purificadas del sistema replicativo de  $\phi$ 29, salvo cuando se especifica lo contrario, fueron: 30 nM de DNA polimerasa y 150 nM TP, incubándose juntas en hielo durante 15 minutos antes de la reacción. Una mezcla de diferentes deoxinucleótidos y dideoxinucleótidos se añadió entonces a la reacción para producir paradas secuenciales y así poder estudiar el proceso de transición. Para el oligonucleótido wt dsDNA de 29 pb se añadieron (i) dATP y ddGTP (dNTP:ddNTP = 5:100  $\mu$ M); (ii) dATP, dGTP y ddTTP (misma relación que en el caso anterior, 5:5:100); (iii) dATP, dGTP, dTTP y ddCTP (idéntica relación, 5:5:5:100); o (iv) 5  $\mu$ M dNTPs cuando se indica, más 1  $\mu$ Ci (13.4 nM)  $[\alpha\text{-}^{32}P]$ dATP. Para los oligonucleótidos A/C y 6m dsDNA de 29 pb se añadieron (mismas relaciones que para el oligonucleótido wt): (i) dGTP y ddTTP; (ii) dGTP, dTTP y ddATP; (iii) dGTP, dTTP, dATP y ddCTP; o (iv) 5  $\mu$ M dNTPs cuando se indica, más 1  $\mu$ Ci (13.4 nM)  $[\alpha\text{-}^{32}P]$ dGTP. Las

reacciones se incubaron durante 5 minutos a 30°C y se detuvieron mediante la adición de 10 mM EDTA y 0.1% SDS, concentraciones finales. Las muestras se procesaron como se ha descrito anteriormente (véase apartado 2.4). En este caso, las muestras se analizaron por SDS-PAGE en geles de poliacrilamida al 12% cuyas dimensiones son 360 x 280 x 0.5 mm, de tal forma que permiten obtener la suficiente resolución como para distinguir todos los productos de elongación unidos a la TP. El número de nucleótidos marcados incorporados por molécula de DNA elongada se tuvo en cuenta para los cálculos en la cuantificación de todas las reacciones. Como marcador de la posición +1, se realizó un experimento de iniciación de la replicación con TP-DNA como molde, además de como forma de estandarización adicional de los experimentos independientes.

### 2.6. Ensayo de replicación con oligonucleótidos dsDNA o ssDNA

La mezcla de reacción contenía, en 25 µl, 50 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 4% (v/v) glicerol, 1 mM DTT, 0.1 mg/ml BSA, 20 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5 µM dNTPs y 1 µCi (13.4 nM) [ $\alpha$ -<sup>32</sup>P]dATP (ratio [ $\alpha$ -<sup>32</sup>P]dATP:dNTPs = 1:373). Las concentraciones de reacción de las proteínas purificadas del sistema replicativo de  $\phi$ 29, salvo cuando se especifica lo contrario, fueron: 30 nM de DNA polimerasa, 150 nM TP, así como 313 nM de los orígenes sintéticos de doble banda especificados. La TP y la DNA polimerasa de  $\phi$ 29 fueron incubadas en hielo durante 15 minutos antes de la reacción. Las reacciones se incubaron 5 minutos a 30°C y se detuvieron mediante la adición de 10 mM EDTA y 0.1% SDS, concentraciones finales. Las muestras se procesaron y analizaron como se ha descrito anteriormente (véase apartado 2.4.).

### 2.7. Transposición y construcción de la librería de inserciones GFP en la TP de $\phi$ 29

El plásmido pGFP-NlaKmTrans contenía en su secuencia un transposón de tipo Tn5 (1819 pb), que consiste en cualquier secuencia flanqueada por unas señales de 19 pb denominadas, en este caso, extremos mosaico (ME, de secuencia 5'CTGTCTCTTATACACATCT3') (Billerbeck *et al.*, 2013). En este caso, el transposón contenía, entre dos secuencias ME, el gen de la GFP sin promotor ni RBS junto al gen de resistencia a kanamicina (Km<sup>R</sup>) regulado por su propio promotor. Con el fin de obtener una cantidad suficiente de transposón para su uso posterior, se realizó la amplificación del fragmento correspondiente del plásmido con los oligonucleótidos Tn5Fw y Tn5Rev. El producto de PCR se purificó con el kit de purificación *QIAquick PCR* (QIAGEN) y posteriormente se incubó con el enzima DpnI durante 4 horas a 37°C con el fin de asegurarse la total eliminación del DNA original. El resultado se volvió a purificar con el mismo kit de purificación.

Para generar la librería de inserción de GFP en la TP del bacteriófago  $\phi$ 29, se realizó una reacción de transposición empleando el transposón amplificado por PCR y el plásmido pT7-TP2ST, que contiene el gen de la TP de  $\phi$ 29 con un doble Strep-tag fusionado en su extremo N-terminal. El gen de la TP está

bajo un promotor T7 que está regulado por IPTG. En un volumen final de 10  $\mu$ l, la reacción de transposición contenía: 50 mM Tris-acetato, pH 7.5, 0.15 M acetato de potasio, 4 mM espermidina, 1 mM EDTA y 10 mM  $Mg_2Cl$ . Se escogió un ratio transposasa/transposón de 10, por lo que se empleó 0.1  $\mu$ M de transposasa y 0.01  $\mu$ M del transposón amplificado por PCR. La mezcla se incubó durante 2 horas a 37°C y se detuvo la reacción con una incubación de 10 minutos a 70°C. Alícuotas de 1  $\mu$ l de la reacción de transposición se emplearon para la transformación de células *E.coli* XL-1 Blue, que fueron posteriormente seleccionadas en base a una doble resistencia kanamicina+ampicilina. Las colonias resistentes se recogieron y se aisló el DNA plasmídico con el kit de purificación *Wizard Plus SV Miniprep* (PROMEGA). La librería resultante se digirió con el enzima de restricción NotI y se volvió a ligar para eliminar el gen de resistencia a kanamicina, obteniendo la librería de TPs fusionadas a GFP. Esta librería se transformó en células *E.coli* BL21(DE3) y las colonias fueron seleccionadas por su resistencia a ampicilina en placas de LB+ampicilina.

## 2.8. Fluorimetría en cultivo líquido

Para el análisis simultáneo de múltiples clones, las colonias bacterianas seleccionadas se cultivaron en 200  $\mu$ l de medio LB-ampicilina durante la noche a 37°C con agitación en placas de 96 pocillos. Tras el crecimiento, la expresión de las proteínas fue inducida durante 2 horas mediante la adición de 0.5 mM IPTG al medio. Una muestra de 50  $\mu$ l de cada pocillo se transfirió a placas negras de 96 pocillos (Greiner Bio-One) y se midió la emisión de fluorescencia en el espectro 485 nm excitación/510 nm emisión, utilizando un equipo lector de placas OPTIMA (BMG LABTECH). En cada placa analizada se destinaron 8 pocillos para cultivos que contenían el plásmido pT7-3 y no expresaban GFP a modo de control negativo (no fluorescentes) y otros 8 para cultivos que contenían el plásmido pEYFP-TP que expresaba la proteína de fusión YFP TP (fusión de YFP al extremo N-terminal de la TP de  $\phi$ 29) a modo de control positivo (fluorescentes).

## 2.9. Análisis de expresión de GFP por Western Blot

Las colonias de *E.coli* BL21(DE3) que contenían la librería de TPs fusionadas a GFP se cultivaron durante la noche a 37°C. Al día siguiente los cultivos se diluyeron 1:100 en medio LB nuevo con el antibiótico correspondiente para restablecer la fase de crecimiento exponencial. A una  $DO_{600}$  de 0.4 la expresión de GFP se indujo añadiendo 0.15 mM IPTG al medio. De cada uno de los cultivos se tomaron alícuotas de 1 ml y fueron centrifugadas a 10000 rpm durante 10 minutos. El sobrenadante se descartó y el *pellet* de células se resuspendió en 250  $\mu$ l de tampón de carga de proteínas (37 mM Tris-HCl, pH 6.8, 2% (p/v) SDS, 4% (v/v) 2-mercaptoetanol, y 13% (v/v) glicerol) para ser sometido posteriormente a sonicación (6 ciclos de 2 segundos a 40 micrones) en un equipo MSE PG1589. Las muestras se calenta-



ron a 95°C durante 5 min y se analizaron por SDS-PAGE en geles de poliacrilamida al 12%. Las proteínas se transfirieron posteriormente a una membrana Immobilon-P de fluoruro de polivinilideno (PVDF) (Millipore) como se describe en Meijer *et al.* (2001b). Las membranas se incubaron con un anticuerpo primario de conejo anti-TP (dilución 1:3000) durante toda la noche a 4°C. La detección de los complejos antígeno-anticuerpo se llevó a cabo mediante la incubación de las membranas con anticuerpos secundarios anti-conejo conjugados con peroxidasa de rábano durante 1 hora a temperatura ambiente y el revelado se realizó con el kit ECL (*Enhanced ChemiLuminescence*) (Amersham) y autorradiografía.

### 2.10. Ensayo de formación del complejo TP-AMP en extractos bacterianos

Las colonias de *E.coli* BL21(DE3) que contenían TPs fusionadas a GFP se cultivaron durante la noche a 37°C. Al día siguiente los cultivos se diluyeron 1:100 en medio LB+ampicilina para restablecer la fase de crecimiento exponencial. A una  $DO_{600}$  de 0.4 la expresión de GFP se indujo añadiendo 0.15 mM IPTG al medio. De cada uno de los cultivos se tomaron alícuotas de 1.5 ml y fueron centrifugadas a 10000 rpm durante 10 min. El sobrenadante se descartó y el *pellet* de células se resuspendió en 300 µl de tampón A que contenía 50 mM Tris-HCl, pH 8, 0.5 M NaCl, 1 mM EDTA, 7 mM 2-mercaptoetanol, 5% (v/v) glicerol y 0.1% Tween-20. La suspensión de células se sonicó como se indica en el *Apartado 2.9*. 1.5 µl del extracto bacteriano obtenido se añadió a una mezcla de reacción que contenía, en 25 µl, 50 mM Tris-HCl, pH 7.5, 25 mM  $MnCl_2$ , 4% (v/v) glicerol, 1 mM DTT, 0.1 mg/ml BSA, 20 mM  $(NH_4)_2SO_4$  y 5 µM de dATP y 1 µCi (13.4 nM)  $[\alpha\text{-}^{32}P]$ dATP, respectivamente (ratio  $[\alpha\text{-}^{32}P]$ dATP:dATPs = 1:373), 35 nM de DNA polimerasa y 0.73 nM  $\phi 29$  TP-DNA. Las reacciones se incubaron 1 hora a 30°C y se detuvieron mediante la adición de 10 mM EDTA y 0.1% (p/v) SDS, concentraciones finales. Las muestras se procesaron como se describe para el ensayo de formación del complejo TP-AMP estándar.

### 2.11. Ensayo de retraso en gel o cambio de movilidad electroforética (EMSA)

El fragmento amplificado por PCR de 216 pb correspondiente al gen *yshC* de *B.subtilis* (Holguera *et al.*, 2014) fue marcado radiactivamente en sus extremos 5' con  $[\gamma\text{-}^{32}P]$ ATP y el enzima polinucleótido quinasa de T4. El nucleótido no incorporado se eliminó mediante el paso de la muestra por columnas *mini Quick Spin DNA columns* (Roche). La mezcla de reacción contenía, en 20 µl, 50 mM Tris-HCl, pH 7.5, 4% (v/v) glicerol, 1 mM DTT, 0.1 mg/ml BSA, 20 mM  $(NH_4)_2SO_4$  y 1 nM del gen *yshC* marcado, así como la concentración indicada de TP empleada en cada caso. Tras una incubación de 5 min a 4°C, las muestras se sometieron a electroforesis en geles de poliacrilamida al 4% con 12 mM Tris-acetato, pH 7.5 y 1 mM EDTA a 25 mA y 4°C durante 1.5 horas. Tras la electroforesis, los geles se secaron y se revelaron por autorradiografía.



### 2.12. Ensayo de competición en iniciación con el TP-DNA de $\phi 29$

El ensayo de competición con TP-DNA se llevó a cabo de forma idéntica al de iniciación de la replicación o formación del complejo TP-dAMP (véase apartado 2.4). En este caso, 75 nM de cada una de las dos TPs a competir se incubaron juntas durante 10 minutos en hielo ( $[TP_{total}] = 150 \text{ nM}$ ). Tras esto, se incubó la mezcla de TPs con 30 nM de DNA polimerasa. Posteriormente se añadieron el resto de componentes de la reacción. Un punto en ausencia de TP se usó como control negativo de actividad. Reacciones con 150 nM de TP wt o TP Strep se llevaron en paralelo como control de la actividad total con cada una de esas proteínas. Como control de la competición, se empleó el mutante de la TP en el residuo iniciador S232C, que es prácticamente inactivo aunque conserva una interacción *wild-type* con la DNA polimerasa (Garmendia *et al.*, 1990), en reacciones junto con la TP wt y la TP Strep. Las muestras se procesaron como se ha descrito para el experimento de formación del complejo TP-dAMP.

### 2.13. Ensayo de amplificación detectado por masa

La mezcla de reacción contenía, en 25  $\mu\text{l}$ , 50 mM Tris-HCl, pH 7.5, 10 mM  $\text{MgCl}_2$ , 4% (v/v) glicerol, 1 mM DTT, 0.1 mg/ml BSA, 20 mM  $(\text{NH}_4)_2\text{SO}_4$  y 100  $\mu\text{M}$  dNTPs. Las concentraciones de reacción de las 4 proteínas purificadas del sistema replicativo de  $\phi 29$ , salvo cuando se especifica lo contrario, fueron: 6 nM de DNA polimerasa, 25 nM de las distintas TP, 23  $\mu\text{M}$  de p5 y 5  $\mu\text{M}$ , en el caso del plásmido pETORPHI, o 30  $\mu\text{M}$ , en el caso del TP-DNA de  $\phi 29$ , de p6, así como las cantidades de DNA molde indicadas en cada experimento. La DNA polimerasa y la TP de  $\phi 29$  fueron incubadas juntas en hielo durante 15 minutos antes de la reacción. Como molde de DNA se empleó 50 nM TP-DNA de  $\phi 29$  o 170 nM de pETORPHI. Las reacciones se incubaron 2 horas a 22°C, en el caso del plásmido pETORPHI, o 30°C, en el caso del TP-DNA de  $\phi 29$ , y se detuvieron mediante la adición de 10 mM EDTA y 0.1% (p/v) SDS, concentraciones finales. Las muestras se analizaron en un gel de agarosa al 0.7% (p/v) que contenía bromuro de etidio y se fotografiaron en un transiluminador. La cuantificación de las bandas se realizó mediante el programa informático ImageJ.

**TPi38cys:** La mezcla de reacción contenía, en 25  $\mu\text{l}$ , 50 mM Tris-HCl, pH 7.5, 10 mM  $\text{MgCl}_2$ , 4% (v/v) glicerol, 1 mM DTT, 0.1 mg/ml BSA, 5 mM  $(\text{NH}_4)_2\text{SO}_4$  y 100  $\mu\text{M}$  dNTPs. Las concentraciones de reacción de las 4 proteínas purificadas del sistema replicativo de  $\phi 29$  fueron: 15 nM DNA polimerasa, 97 nM de TPi38cys, 23  $\mu\text{M}$  p5 y 0.8  $\mu\text{M}$  p6. Como molde de DNA se empleó 170 nM de pEYFPORBae. La DNA polimerasa y la TPi38cys fueron incubadas juntas en hielo durante 15 minutos antes de la reacción. Las reacciones se incubaron 2 horas a 22°C y se detuvieron mediante la adición de 10 mM EDTA y 0.1% (p/v) SDS, concentraciones finales. Las muestras se analizaron en un gel de agarosa al 0.7% (p/v) que contenía bromuro de etidio y se fotografiaron en un transiluminador.

### 2.14. Conjugación de la proteína TPi38cys con el grupo maleimida

Las cantidades indicadas de proteína TPi38cys se incubaron con un exceso molar de 20 veces los péptidos Mal-TAT o Mal-ANTP en tampón 50 mM HEPES pH 7.5 durante toda la noche a 4°C. Las reacciones de conjugación se analizaron por SDS-PAGE en geles de poliacrilamida al 12% y posterior tinción por azul de Coomassie.

## Resultados




# Capítulo 1

**Método de amplificación de DNAs heterólogos basado en el sistema replicativo del bacteriófago  $\phi$ 29**

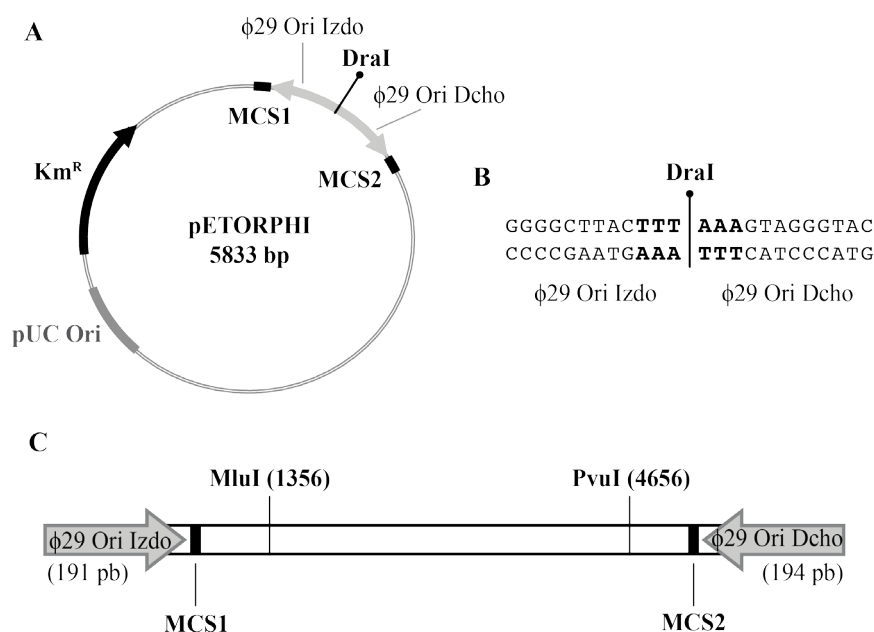


Algunos sistemas de amplificación isotérmica han surgido en respuesta a las limitaciones de la técnica de PCR en la amplificación de ácidos nucleicos. En algunos casos, el complejo diseño de los cebadores o la necesidad del procesamiento posterior del producto aún suponen obstáculos en el uso de estos métodos. El sistema replicativo de  $\phi 29$  ha demostrado ser una potente herramienta de amplificación de ácidos nucleicos (Blanco *et al.*, 1994) que ha servido como plataforma de sistemas de amplificación isotérmica como la SDA o la MDA sin la necesidad del diseño de cebadores. Sin embargo, hasta el momento no se ha desarrollado un sistema de amplificación basado en  $\phi 29$  que, empleando una de sus características distintivas, produjera DNA con TP unida covalentemente a sus extremos 5' a partir de un DNA heterólogo, lo que supondría potenciales ventajas y posibilitaría aplicaciones que serán discutidas más adelante. Con el objeto de desarrollar este método de amplificación, se diseñó un plásmido cuya linealización producía un DNA flanqueado por los orígenes de replicación del genoma de  $\phi 29$ .

### 1. Diseño de un plásmido cuya linealización produce un DNA flanqueado por los extremos funcionales del genoma de $\phi 29$

El análisis de la eficiencia de iniciación en moldes que contenían los orígenes de replicación de  $\phi 29$  sin TP mostró que la reacción era mucho menos eficiente (unas 30 veces menos) que con el TP-DNA de  $\phi 29$  como sustrato (Gutiérrez *et al.*, 1986). Sin embargo, hasta el momento no se había puesto a punto un sistema de amplificación *in vitro* de DNAs heterólogos basado en el sistema replicativo *in vitro* de  $\phi 29$  (Blanco *et al.*, 1994). Para que un DNA se constituya como sustrato amplificable por la maquinaria replicativa de  $\phi 29$  requiere de los dos extremos funcionales del genoma correctamente orientados. De esta forma, para observar si era posible obtener amplificación utilizando como molde un DNA flanqueado por los orígenes de replicación de  $\phi 29$  sin TP unida en los extremos 5', se diseñó el plásmido pETORPHI (5.8 Kb) (**Figura 8A**). Este plásmido contiene los 191 y 194 pb de los extremos izquierdo y derecho del genoma, respectivamente, un origen de replicación de tipo pUC (pUC Ori), un marcador de resistencia a kanamicina ( $Km^R$ ) y dos sitios de policlonaje (MCS1 y MCS2). Para garantizar la correcta orientación de los extremos del genoma de  $\phi 29$  se clonaron unidos y en sentido inverso. Esto genera un sitio de reconocimiento único en el plásmido para el enzima de restricción *DraI* y mediante su corte, se obtiene un DNA lineal flanqueado por los extremos del genoma de  $\phi 29$  (**Figura 8B**). La longitud de la secuencia de los extremos clonada de 191 y 194 pb para los extremos izquierdo y derecho, respectivamente, se determinó en base al sitio de reconocimiento de la proteína p6 por su papel fundamental en la amplificación del DNA de  $\phi 29$ . Como se ha mencionado en la Introducción, la proteína p6 no parece reconocer una secuencia determinada sino una característica estructural como un patrón de flexibilidad en la hélice de DNA, habiendo

sidio determinado su sitio de nucleación entre las posiciones 62-125 y 46-68 para los extremos derecho e izquierdo del genoma de  $\phi 29$ , respectivamente. A pesar de que el *footprint* obtenido con p6 muestra que su unión se extiende más allá, a partir de la posición 200 ésta se torna más débil, especialmente a altas concentraciones de sal (Serrano *et al.*, 1989). Por esa razón se clonaron los primeros 191 y 194 pb de los extremos izquierdo y derecho, respectivamente, con una secuencia idéntica al genoma (**Figura 8C**). Como se ha mencionado anteriormente, estos fragmentos contenían, además, los primeros 12 pb de cada extremo, caracterizados como los orígenes mínimos de replicación (Gutiérrez *et al.*, 1988), que incluyen las ITRs de 6 pb (5' AAAGTA<sup>3'</sup>) y, a su vez, el triplete AAA en el extremo necesario para un correcto mecanismo de *sliding-back* en la iniciación de la replicación (Méndez *et al.*, 1992). Estos fragmentos constituyen lo que se denominó orígenes de amplificación.



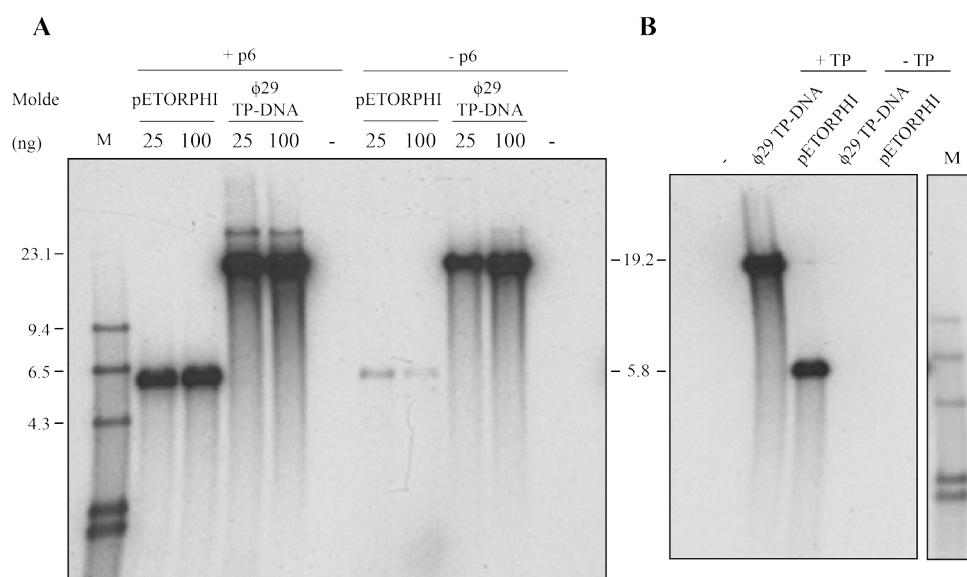
**Figura 8. Diseño del plásmido pETORPHI.** (A) El plásmido pETORPHI contiene clonados los fragmentos correspondientes a los 191 y 194 pb de los extremos izquierdo (Ori Izdo) y derecho (Ori Dcho), respectivamente, del genoma del bacteriófago  $\phi 29$ . Además, contiene un marcador de resistencia a kanamicina (Km<sup>R</sup>), un origen de tipo pUC (pUC Ori) y dos sitios de policlonaje (MCS1 y MCS2). Las secuencias de los sitios de policlonaje se detallan en la **Tabla S1**. (B) Los extremos se clonaron juntos y en una orientación inversa, de forma que se generaba una diana para el enzima de restricción DraI. (C) La linealización mediante el corte con DraI produce un DNA lineal flanqueado por los extremos del genoma de  $\phi 29$  en una orientación correcta.



## 2. Amplificación *in vitro* por la maquinaria replicativa del bacteriófago $\phi 29$ del plásmido pETORPHI linealizado

En primer lugar se realizaron una serie de experimentos para determinar las condiciones experimentales óptimas de amplificación (no mostrado). Estas condiciones resultaron ser, en un volumen final de 25  $\mu$ l: 50 mM Tris-HCl pH 7.5, 10 mM  $MgCl_2$ , 5% glycerol, 1 mM DTT, 0.1 mg/mL BSA y 20 mM sulfato amónico. En cuanto al sistema de amplificación basado en el bacteriófago  $\phi 29$ , las concentraciones óptimas en la reacción de las 4 proteínas purificadas resultaron: 25 nM de TP, 6 nM de DNA polimerasa, 23  $\mu$ M de p5 y 5  $\mu$ M de p6. La temperatura óptima de reacción resultó ser 22 °C.

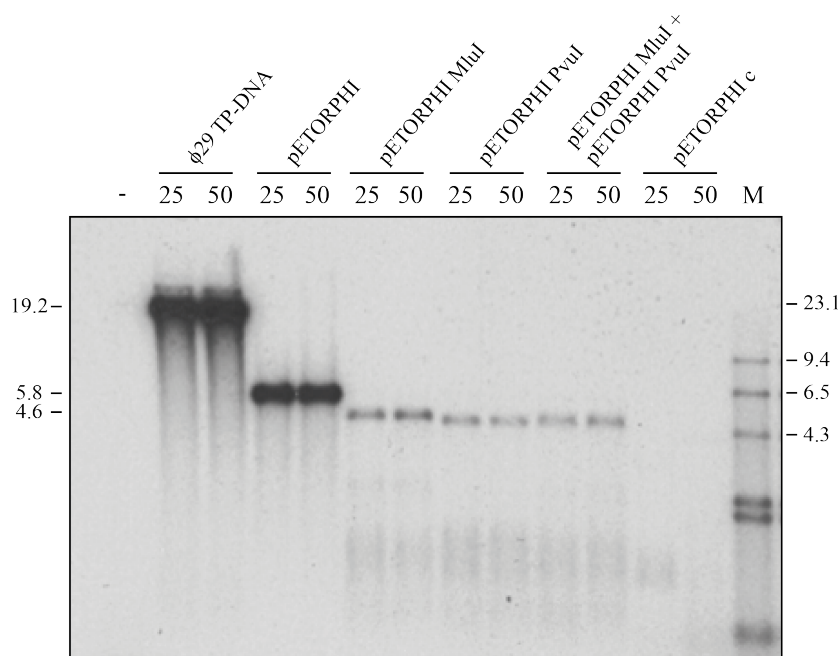
En la **Figura 9A** se muestra el resultado de un experimento en el que se observa cómo el pETORPHI cortado con el enzima DraI (lineal) era capaz de ser utilizado como sustrato amplificable por la maquinaria de replicación de  $\phi 29$ . La actividad se midió por incorporación de material radiactivo en el DNA de nueva síntesis. El experimento de amplificación se realizó en presencia y ausencia de la proteína p6, lo que permitía observar si el sitio de nucleación para p6 comprendido en los primeros 200 pb era funcional en el caso de un DNA heterólogo.



**Figura 9. Producción de DNA con el sistema replicativo de  $\phi 29$ .** El experimento de amplificación se realizó como se describe en *Materiales y Métodos*. Las reacciones se incubaron durante 2 horas a 22°C. Como marcador de peso molecular se empleó el DNA del fago  $\lambda$  cortado con el enzima HindIII (carril M). **(A)** Se empleó 260 pM (25 ng) y 1 nM (100 ng) de plásmido pETORPHI y 78 pM (25 ng), y 310 pM (100 ng) de TP-DNA de  $\phi 29$ . Se indica la presencia (carriles +p6) o ausencia (carriles -p6) de proteína p6, así como el tamaño de los productos (19.2 Kb para el TP-DNA de  $\phi 29$  y 5.8 Kb para el pETORPHI linealizado con DraI). **(B)** Se empleó 260 pM de plásmido pETORPHI y 78 pM de TP-DNA de  $\phi 29$  (25 ng en ambos casos) en presencia (carril +TP) o ausencia (carril -TP) de TP. Las figuras son representativas de, al menos, tres experimentos independientes.

En presencia de p6, con el TP-DNA de  $\phi 29$  se alcanzó una cantidad de producto de tamaño correcto (19.2 Kb) similar, independientemente de la cantidad de molde inicial (25 o 100 ng), lo que suponía que, en las condiciones de reacción empleadas, el nivel de producto había alcanzado una meseta con 25 ng. Con el plásmido pETORPHI linealizado como molde también se observó amplificación y el producto de tamaño correcto (5.8 Kb) también había alcanzado una meseta con 25 ng.

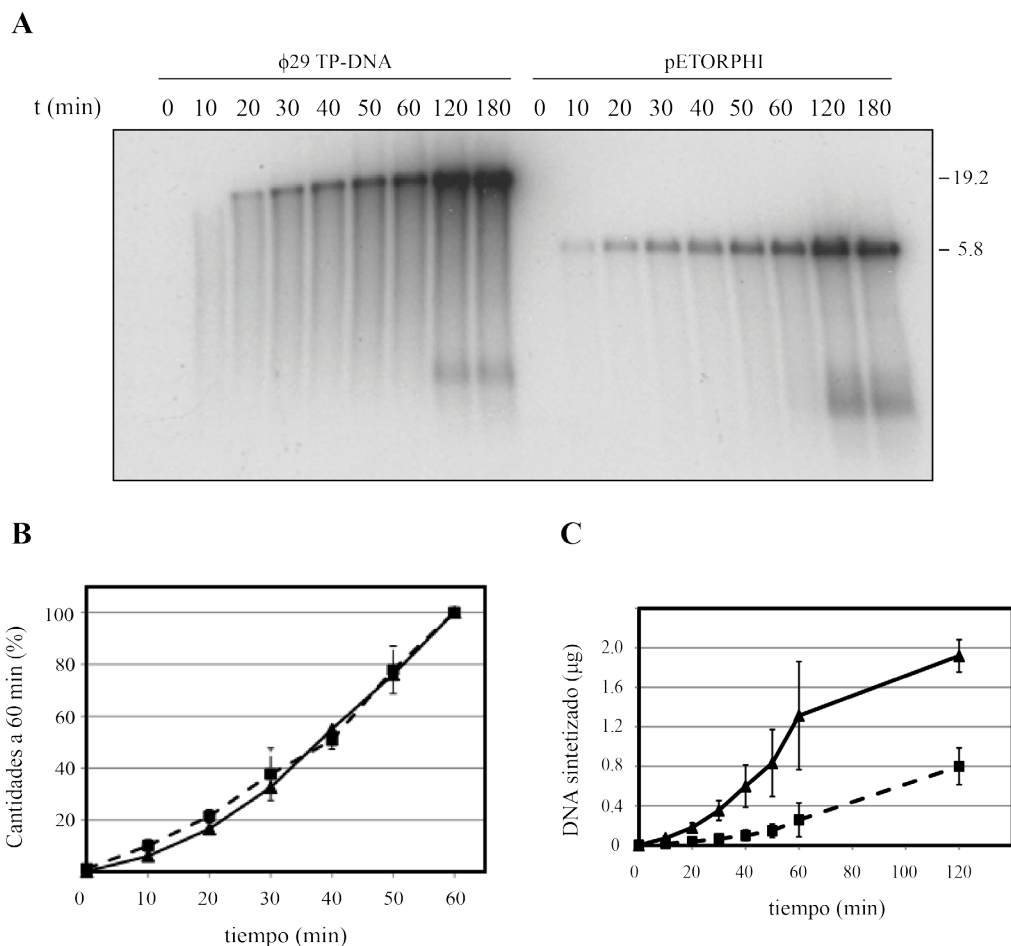
En ausencia de p6, también se observó síntesis de TP-DNA de  $\phi 29$ , pero ésta no alcanzaba el nivel obtenido en presencia de p6 ni el punto de saturación de producto con 25 ng. En el caso del pETORPHI, la síntesis fue muy pobre. Esto sugiere que la síntesis de DNA con el pETORPHI, en mayor medida que con el TP-DNA de  $\phi 29$ , es fuertemente estimulada por p6.



**Figura 10. Amplificación de DNA con el sistema replicativo de  $\phi 29$ .** Las digestiones y los experimentos de amplificación se realizaron como se describe en *Materiales y Métodos*. Las reacciones se incubaron durante 2 horas a 22°C. Se empleó 78 pM (25 ng) y 155 pM (50 ng) de TP-DNA de  $\phi 29$  (carril  $\phi 29$  TP-DNA), así como 260 pM (25 ng) y 520 pM (50 ng) de plásmido pETORPHI (carril pETORPHI), plásmido pETORPHI digerido con MluI (carril pETORPHI MluI), plásmido pETORPHI digerido con PvuI (carril pETORPHI PvuI), una mezcla del producto de la digestión de pETORPHI con MluI y el producto de la digestión con PvuI (carril pETORPHI MluI + pETORPHI PvuI), y plásmido pETORPHI circular sin linealizar (carril pETORPHI c). Como marcador de peso molecular se empleó el DNA del fago  $\lambda$  cortado con el enzima HindIII (carril M). El tamaño de las bandas correspondientes al producto de la digestión del plásmido linealizado con MluI o PvuI son 4.6 y 4.4 Kb, respectivamente. La figura es representativa de, al menos, tres experimentos independientes.

Para asegurar que la síntesis de DNA estaba siendo el resultado de un mecanismo cebado por TP, se realizó como control un experimento de amplificación con el TP-DNA de  $\phi 29$  y el pETORPHI linealizado en presencia y ausencia de TP. Como se observa en la **Figura 9B**, la síntesis de DNA con ambos moldes era completamente dependiente de la presencia de TP, lo que supone que la producción de DNA de nueva síntesis correspondía a eventos de iniciación con TP y que el material sintetizado contiene TP unida covalentemente a ambos extremos 5'.

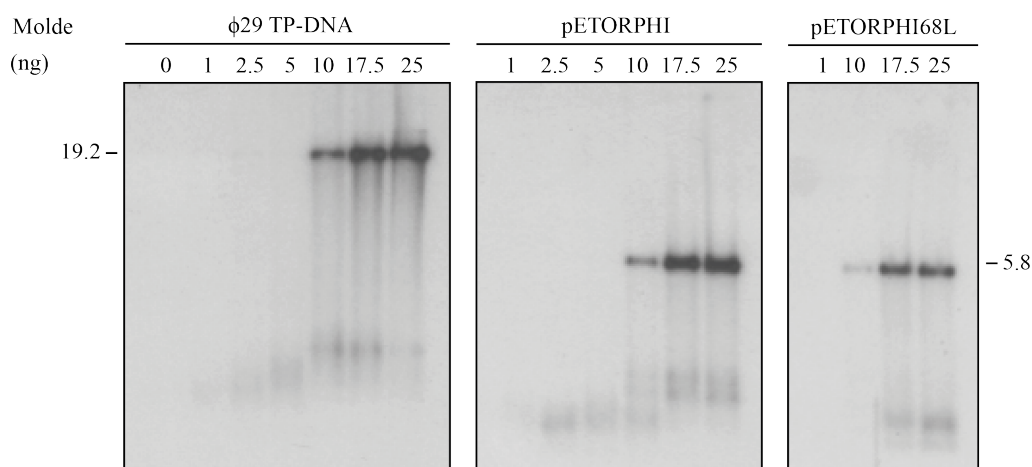
Para comprobar que la producción de DNA empleando el plásmido pETORPHI linealizado como molde era consecuencia de una amplificación real y no de eventos de replicación unidireccionales (copia de una única hebra), se realizó un experimento de amplificación con el plásmido pETORPHI linealizado con DraI y además cortado con los enzimas MluI o PvuI, que tienen un sitio de reconocimiento único en el plásmido (**Figura 10**). Estos fragmentos no pueden ser amplificados puesto que cada uno contiene uno de los orígenes de amplificación y no hay continuidad en la doble cadena de DNA entre ambos, por lo que solo son posibles eventos de replicación lineal de ambos fragmentos iniciados desde un extremo. En la **Figura 10** se observa cómo en estos casos (carriles pETORPHI MluI y pETORPHI PvuI) la cantidad de DNA sintetizado decreció sustancialmente (entre 10-20 veces, dependiendo del caso) en comparación con el producto obtenido con el pETORPHI linealizado con DraI (carril pETORPHI). Para descartar posibles hibridaciones o complementación entre fragmentos, los productos de ambas digestiones se mezclaron en una misma reacción de amplificación (carril pETORPHI MluI + pETORPHI PvuI), obteniendo una producción de DNA similar a los casos por separado. Como se esperaba, en el control del plásmido pETORPHI circular no cortado con DraI (carril pETORPHIc) no se obtuvo actividad, ya que los orígenes de replicación no eran accesibles a la maquinaria replicativa. Estos resultados muestran que el molde pETORPHI linealizado produce una amplificación real basada en el sistema de  $\phi 29$  y mediada por el uso de los orígenes de amplificación en los extremos de la molécula. Por otra parte, se compararon las cinéticas de amplificación del TP-DNA de  $\phi 29$  y pETORPHI. La **Figura 11A** muestra que con los dos moldes se obtiene un patrón similar de amplificación, llegando a la producción máxima a los 120 minutos. Si se normaliza la cantidad de DNA sintetizada tomando los valores obtenidos a los 60 minutos como el 100%, ambos moldes, tanto el TP-DNA de  $\phi 29$  como el pETORPHI, siguen un patrón de amplificación no lineal prácticamente idéntico (**Figura 11B**). Esto parece apoyar los resultados anteriores en el hecho de que el pETORPHI linealizado produce una amplificación similar a la obtenida con el TP-DNA de  $\phi 29$  (Blanco *et al.*, 1994). Partiendo de 25 ng de molde, el factor de amplificación resultó de  $83 \pm 19$  DE para el TP-DNA de  $\phi 29$  y de  $30 \pm 7$  DE para el pETORPHI linealizado, que corresponde aproximadamente a 2 o 0.8  $\mu$ g, respectivamente, de DNA total sintetizado (**Figura 11C**).



**Figura 11. Cinética de amplificación del TP-DNA de  $\phi 29$  y el plásmido pETORPHI linealizado.** (A) El experimento de amplificación se realizó como se describe en *Materiales y Métodos*. Las reacciones contenían 78 pM (25 ng) de TP-DNA de  $\phi 29$  o 260 pM (25 ng) de plásmido pETORPHI y se incubaron durante a 22°C durante los tiempos indicados. Se indica el tamaño de los productos (19.2 Kb para el TP-DNA de  $\phi 29$  y 5.8 Kb para el pETORPHI cortado con DraI). (B) La gráfica muestra las cantidades de DNA de nueva síntesis, expresadas como el porcentaje del valor obtenido a 60 min como 100%, representadas en función del tiempo para el molde TP-DNA de  $\phi 29$  ( $\blacktriangle$ ) y pETORPHI linealizado ( $\blacksquare$ ). (C) La gráfica muestra la cantidad de DNA sintetizado ( $\mu$ g) en función del tiempo para el molde TP-DNA de  $\phi 29$  ( $\blacktriangle$ ) y pETORPHI linealizado ( $\blacksquare$ ). Los datos presentan los valores medios y la desviación estándar, obtenidos a partir de los datos de tres experimentos independientes.

### 3. Determinación de la cantidad mínima de molde para la amplificación del plásmido pETORPHI linealizado

Experimentos anteriores mostraron que en presencia de la proteína p6, con 25 ng de DNA molde se había alcanzado una meseta en amplificación (**Figura 9A**), por lo que se trató de determinar la cantidad mínima de DNA molde que permitía la reacción de amplificación con el pETORPHI. Para ello se realizó un experimento de amplificación con cantidades menores de molde inicial (**Figura 12**). En las condiciones empleadas, la mínima cantidad requerida de TP-DNA de  $\phi 29$  fue de 10 ng, alcanzándose con 17.5 ng un valor muy cercano a la saturación. En el caso del pETORPHI, los resultados fueron similares, con un mínimo de 10 ng y próximo al valor de saturación con 17.5 ng. Cabe destacar que la diferencia de tamaño entre ambos sustratos hace que la misma masa no se corresponda con el mismo número de moléculas, de forma que para una masa determinada hay 3.3 veces más moléculas de pETORPHI que de TP-DNA de  $\phi 29$ . Como se describió previamente, la secuencia mínima del genoma de  $\phi 29$  que permite la replicación estimulada por p6 corresponde a los primeros 68 pb en el caso del extremo izquierdo, o a los primeros 125 pb del extremo derecho, aunque en este caso el sitio de alta afinidad de p6 no está tan bien definido como en el primero (Serrano *et al.*, 1989, Serrano *et al.*, 1993).



**Figura 12. Cantidad mínima de molde requerida para la amplificación.** El experimento de amplificación se realizó como se describe en *Materiales y Métodos*. Se emplearon las cantidades de plásmido pETORPHI y de TP-DNA de  $\phi 29$  indicadas en la figura. El plásmido pETORPHI68L, al que se le sustituyeron las 194 pb del origen mínimo de amplificación del extremo derecho de  $\phi 29$  por los primeros 68 pb del extremo izquierdo, se construyó como se describe en *Materiales y Métodos*. La figura es representativa de, al menos, tres experimentos independientes.

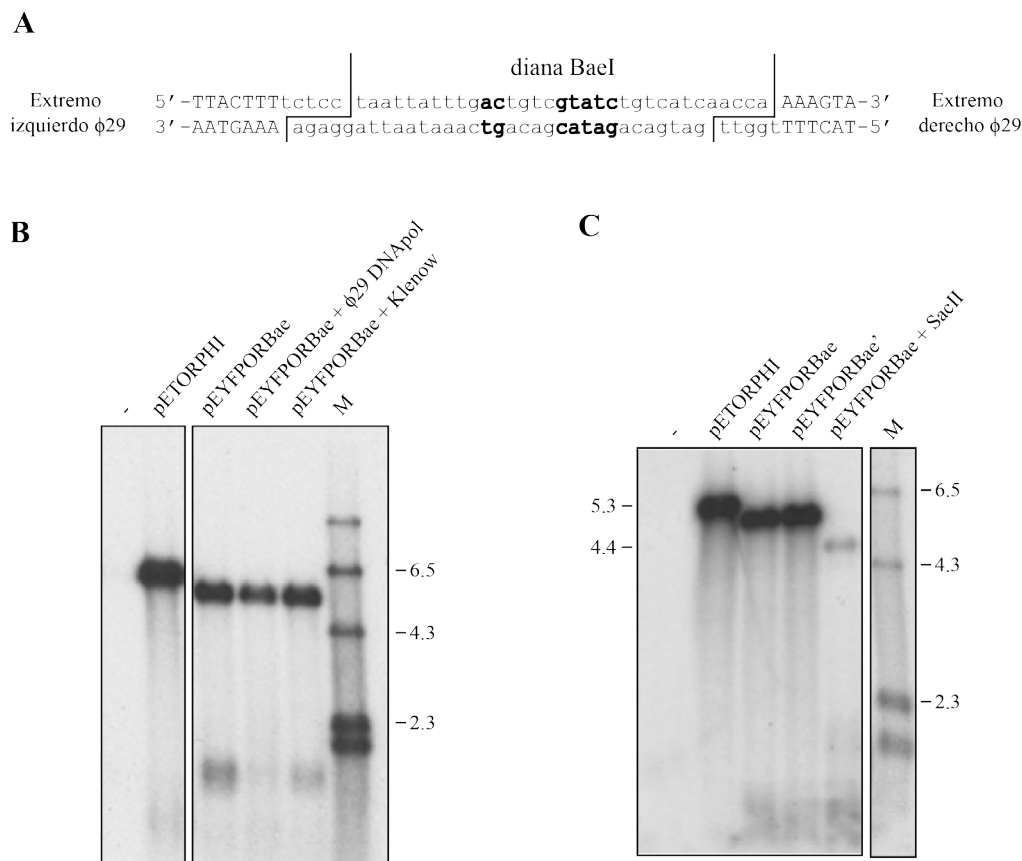
Para desarrollar un sistema de amplificación basado en secuencias mínimas, se comprobó si los 68 pb del extremo izquierdo de  $\phi 29$  eran capaces de producir amplificación. De esa forma, el segmento de 194 pb correspondiente al extremo derecho de  $\phi 29$  que contenía el plásmido pETORPHI fue sustituido por la secuencia mínima de 68 pb del extremo izquierdo para obtener el plásmido pETORPHI68L. El fragmento de 68 pb fue clonado de forma análoga al clonaje del extremo derecho para garantizar su correcta orientación tras el corte con el enzima DraI. Como se muestra en la **Figura 12**, la cantidad mínima de pETORPHI68L que sirve como molde en el sistema de  $\phi 29$  es la misma que la del pETORPHI original, aunque la cantidad final de producto obtenido es menor (un 30% la actividad obtenida con el pETORPHI linealizado). Esta disminución en el rendimiento de la reacción podría ser consecuencia de que la secuencia de 68 pb no tiene la misma eficiencia en la unión de p6 que los 194 pb del pETORPHI.

#### 4. Otras estrategias para producir TP-DNAs heterólogos

La secuencia diana de reconocimiento para el enzima DraI, es muy frecuente en vectores y genomas con un contenido medio de AT, por lo que la linealización del plásmido pETORPHI podría suponer un problema en aquellos casos en los que el DNA flanqueado por los orígenes de amplificación de  $\phi 29$  tuviera un contenido medio o alto en AT. Para solucionar este problema, se procedió al desarrollo de dos estrategias adicionales para producir TP-DNAs heterólogos con los orígenes de replicación del DNA de  $\phi 29$ .

En primer lugar, se construyó un plásmido similar al pETORPHI, es decir, con los orígenes de  $\phi 29$ , pero al que se le había sustituido la diana de reconocimiento de DraI por la de BaeI y que contenía el gen de la YFP al que se denominó pEYFPORBae (véase *Materiales y Métodos*) (**Figura 13A**). La secuencia de reconocimiento del enzima BaeI es compleja y, por tanto, no es tan frecuente en los DNAs como la de DraI. El enzima BaeI corta en dos puntos diferentes las dos hebras del DNA, puntos separados por una distancia de 33 nucleótidos. La secuencia de reconocimiento se encuentra en el medio del fragmento liberado, produciéndose el corte a ambos lados, independientemente de la secuencia y dejando extensiones de banda simple en 3' de 5 nucleótidos (**Figura 13A**). Puesto que no se conocía el efecto de extremos 3' protuberantes en la utilización de los orígenes de replicación de  $\phi 29$ , se llevaron a cabo tres aproximaciones para la reacción de amplificación (**Figura 13B**): (i) una reacción de amplificación estándar tras el corte con BaeI (carril pEYFPORBae); (ii) una incubación con la DNA polimerasa de  $\phi 29$  del plásmido digerido con BaeI en presencia de dNTPs para eliminar las extensiones 3', dejando extremos romos, y posterior adición de la TP y el resto de componentes de la reacción (carril pEYFPORBae +  $\phi 29$  DNApol); y (iii) una incubación con la polimerasa Klenow del plásmido digerido con BaeI en presencia de dNTPs



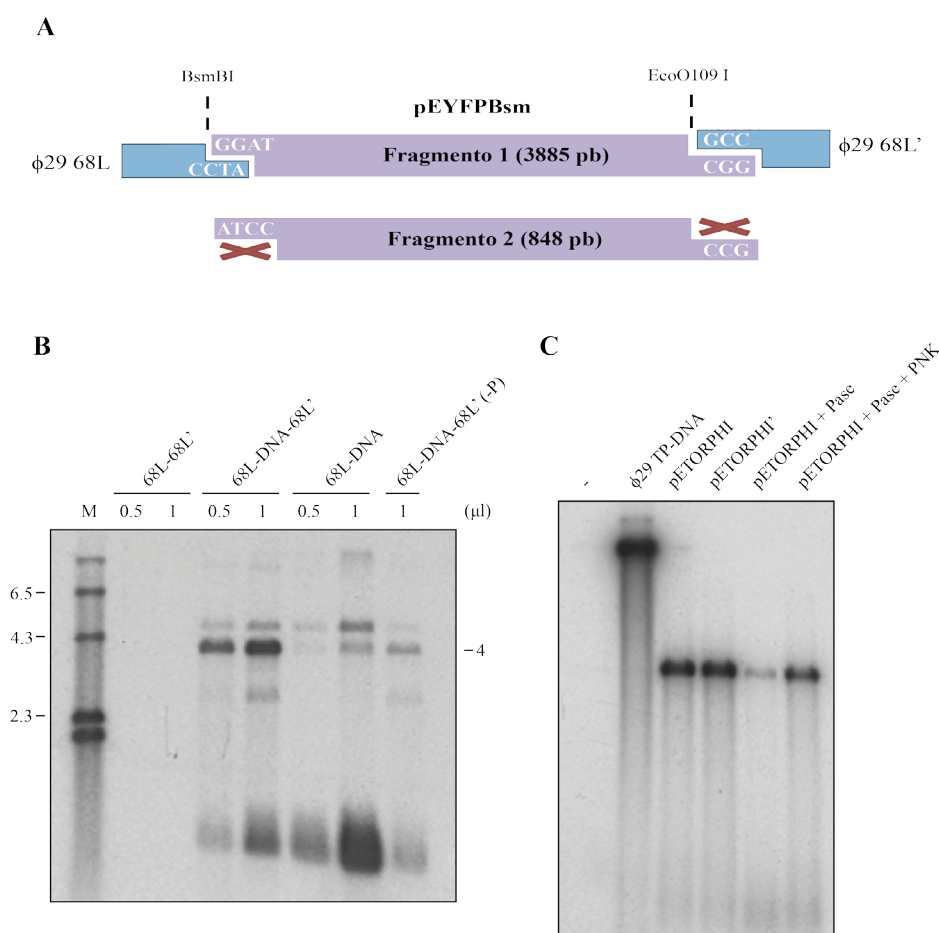


**Figura 13. Diseño y amplificación del plásmido pEYFPORBae.** El plásmido pEYFPORBae se construyó como se especifica en *Materiales y Métodos*. El experimento de amplificación se realizó como se describe en *Materiales y Métodos*. **(A)** Esquema de la secuencia de reconocimiento (negrita) y sitios de corte del enzima de restricción BaeI en el plásmido pEYFPORBae. Los extremos del genoma de  $\phi 29$  se representan con letras mayúsculas. **(B)** Se empleó 285 pM (25 ng) de pEYFPORBae en una reacción de amplificación estándar (carril pEYFPORBae) y se compararon con la amplificación de otros plásmidos que se habían sometido a dos tratamientos previos diferentes: (i) una incubación con la DNA polimerasa de  $\phi 29$  y 100  $\mu$ M dNTPs durante 10 min y posterior adición de la TP (carril pEYFPORBae +  $\phi 29$  DNAPol) y (ii) una incubación previa con el enzima Klenow y 30  $\mu$ M dNTPs durante 10 min, posterior inactivación del enzima y adición del resto de componentes de la reacción (carril pEYFPORBae + Klenow). Como control de amplificación se empleó 260 pM (25 ng) de plásmido pETORPHI (carril pETORPHI). El tamaño de los productos corresponde a 5.8 Kb para el pETORPHI cortado con DraI y a 5.3 Kb para el pEYFPORBae cortado con BaeI. **(C)** Se realizó un experimento de amplificación con el plásmido pEYFPORBae cortado con BaeI (carril pEYFPORBae) y con un pEYFPORBae linealizado e incubado sin (carril pEYFPORBae') y con SacII (pEYFPORBae + SacII). El corte adicional producido por el enzima SacII en el plásmido pEYFPORBae linealizado con BaeI genera dos fragmentos de 4.4 y 0.9 Kb.

para generar también extremos romos y posterior adición de la TP y demás componentes de la reacción (carril pEYFPORBae + Klenow). Como se muestra en la **Figura 13B**, la producción de DNA de las aproximaciones i y iii fue un 50% de la obtenida con el pETORPHI original. El tratamiento con la DNA polimerasa de  $\phi 29$  produjo una cantidad menor de DNA, probablemente como resultado de la fuerte actividad exonucleolítica del enzima. Como control de amplificación genuina basada en el sistema de  $\phi 29$ , un plásmido pEYFPORBae digerido con BaeI fue tratado adicionalmente con el enzima de restricción SacII para interrumpir la cadena de DNA entre los orígenes, resultando una cantidad muy baja de DNA asociado a eventos de replicación unidireccionales como se ha sugerido previamente (**Figura 13C**).

Como se demostró anteriormente, un fragmento de 68 pb con la secuencia del extremo izquierdo del genoma de  $\phi 29$ , si bien con menor actividad, sigue siendo un origen de amplificación funcional. De esta forma, fragmentos de dicha extensión y secuencia pueden ser diseñados en forma de adaptadores con secuencias adicionales, como extremos cohesivos que permitan la ligación a diferentes extremos de DNA obtenidos tras la digestión con un enzima de restricción. Para probar esto, se introdujo una diana de restricción para BsmBI mediante mutagénesis dirigida en el plásmido pEYFP-N1 en la posición 4702 (plásmido pEYFPBsm), y se digirió el plásmido con los enzimas BsmBI y Eco0109I, ambos enzimas con secuencias de reconocimiento únicas en el vector. Se diseñaron dos adaptadores con los primeros 68 pb del extremo izquierdo del genoma de  $\phi 29$  y las extensiones en banda simple específicas para ser ligadas a los extremos resultantes de la digestión del plásmido pEYFPBsm con los enzimas BsmBI y Eco0109I (**Figura 14A**). Estos enzimas se escogieron por generar extremos asimétricos, de forma que los adaptadores no pueden ligar uno con otro ni consigo mismos. Del mismo modo, cada adaptador se une específicamente a uno de los extremos del fragmento mayor resultante de la digestión del plásmido. Este abordaje hace innecesaria la purificación por gel de los productos de ligación previa a la reacción de amplificación, con la correspondiente pérdida de material amplificable, ya que sólo el producto de 4021 pb resultante de la reacción de ligación (el fragmento del plásmido unido a los dos adaptadores) es el único que puede ser molde para una amplificación como se ha demostrado anteriormente. El otro fragmento resultante de la digestión del plásmido (fragmento 2848 pb) no puede ligarse a los adaptadores y, por tanto, no puede ser amplificado. Como se observa en el **Figura 14B**, sólo la ligación que contiene los dos extremos genera una banda de amplificación en la posición correcta (carril 68L-DNA-68L'), mientras que en la reacción en que se ligó un solo adaptador se observó una señal mucho más baja (carril 68L-DNA). El control que contenía únicamente los adaptadores no dio ningún producto (carril 68L-68L'). Dado que todos los moldes que se habían empleado anteriormente tenían un grupo fosfato en el extremo 5', la **Figura 14B** también muestra la importancia de este fosfato en 5' ya que cuando se emplearon los mismos adaptadores sin él (carril 68L-DNA-68L' -P) se obtuvo una muy baja actividad. Las bandas de alto peso molecular (aproximadamente 4.7 Kb) que aparecen por encima de los productos con el tamaño correcto podrían corresponder al plásmi-





**Figura 14. Diseño de adaptadores de amplificación basados en los 68 primeros pb del genoma de  $\phi 29$ .**

(A) El plásmido pEYFPBsm se construyó como se especifica en *Materiales y Métodos*. Su digestión con las enzimas BsmBI y EcoO109I produce los fragmentos 1 y 2 indicados. Los adaptadores de 68 pb se construyeron mediante la hibridación de oligonucleótidos de banda simple, de forma que se generaban extremos cohesivos para los cortes con las enzimas BsmBI (68L) y EcoO109I (68L') que solo permitían su ligación con el fragmento 1. (B) Se realizó un experimento de amplificación como se describe en *Materiales y Métodos* con alícuotas de 0.5 o 1  $\mu$ l de las siguientes ligaciones: (i) ligación del plásmido pEYFP-N1Bsm con uno de los adaptadores (carril 68L-DNA), (ii) ligación del plásmido pEYFP-N1Bsm con los dos adaptadores (carril 68L-DNA-68L') y (iii) ligación de los dos adaptadores (carril 68L-68L'). Una alícuota de la muestra 68L-DNA-68L' fue tratada con la Fosfatasa Antártica (NEB) para eliminar el grupo fosfato en 5' (carril 68L-DNA-68L' (-P)). (C) Alícuotas que contenían 260 pM (25 ng) de plásmido pETORPHI linealizado fueron (i) no tratadas (carril pETORPHI), (ii) tratadas con Fosfatasa Antártica (carril pETORPHI + Pase), (iii) tratadas con Fosfatasa Antártica y posterior tratamiento con PNK (carril pETORPHI + Pase + PNK) o (iv) procesadas de la misma forma que las muestras que tenían un tratamiento pero sin la adición de los enzimas (carril pETOTPHI'), y fueron sometidas a un experimento de amplificación como se describe en *Materiales y Métodos*. El tamaño del pETORPHI linealizado es de 5.8 Kb y el del producto amplificable de la muestra 68L-DNA-68L' es 4.0 Kb.

do digerido con uno solo de los enzimas y ligado únicamente a uno de los adaptadores, siendo sustratos no válidos para la amplificación. Las bandas de bajo peso molecular que aparecen, especialmente en el molde con un solo adaptador (68L-DNA) podrían corresponder a productos que han sido caracterizados como moléculas truncadas que se vuelven palindrómicas y adquieren un origen doble de replicación (Esteban *et al.*, 1997). Los resultados obtenidos mostraron que era posible la ligación de orígenes de amplificación mínimos de  $\phi 29$  a DNAs heterólogos para obtener moléculas de TP-DNA de un tamaño correcto mediante su amplificación con el sistema replicativo de  $\phi 29$ .

La importancia del grupo fosfato en posición 5' se estudió en el experimento mostrado en la **Figura 14C**. Cuando el plásmido pETORPHI se linealiza con el enzima *DraI* y es tratado posteriormente con Fosfatasa Antártica (NEB) se obtiene una síntesis de DNA muy baja (carril pETORPHI +Pasa). Un posterior tratamiento con la Polinucleótido Quinasa de T4 (carril pETORPHI +Pasa +PNK) recuperaba la eficiencia de amplificación hasta niveles próximos a los obtenidos con el pETORPHI sin tratar (carril pETORPHI) o con el tratado como control en las mismas condiciones que los anteriores puntos (carril pETORPHI'). La importancia del fosfato en posición 5' podría deberse a que en la iniciación cebada por TP, el grupo fosfato se acomoda mejor en el bolsillo formado por el heterodímero TP/DNA polimerasa y se establecen una serie de interacciones que estabilizan al complejo ternario.

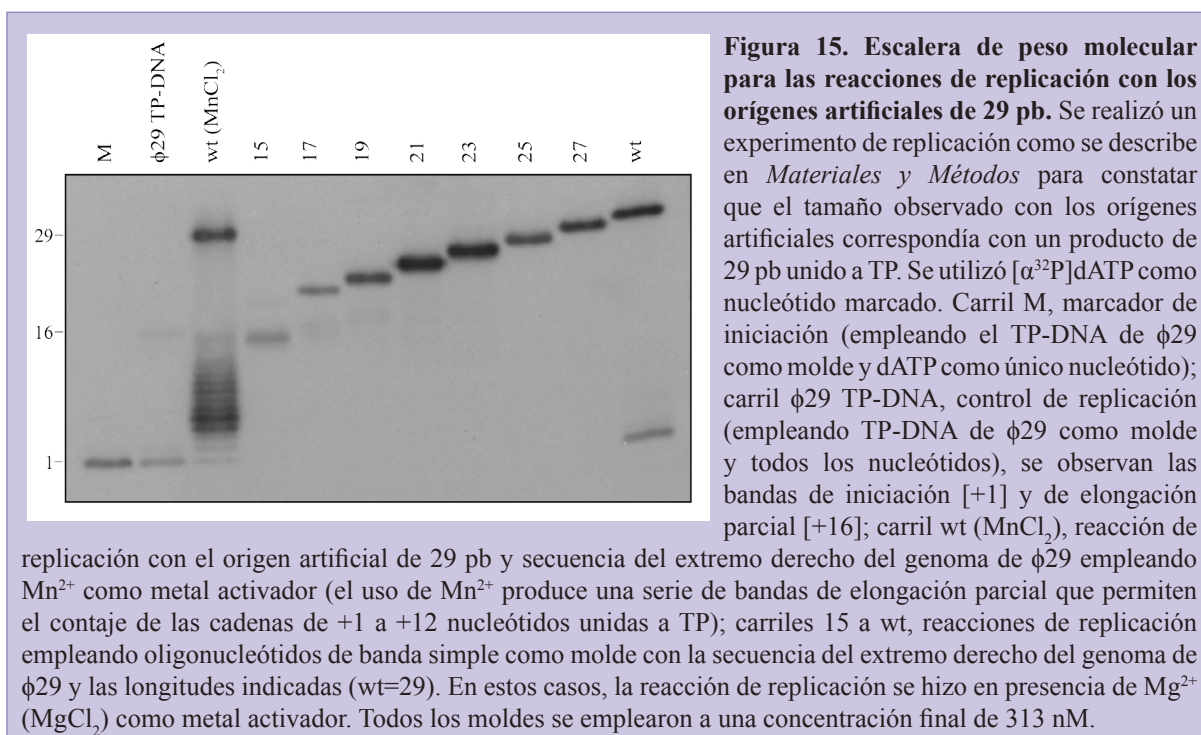
## Capítulo 2

**Mejora de los orígenes de replicación y estudio de las primeras fases del proceso replicativo del bacteriófago  $\phi 29$**



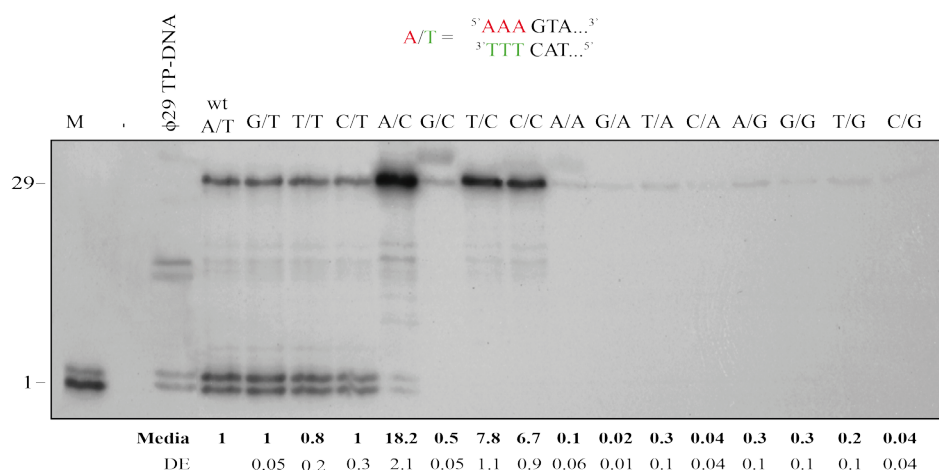
El sistema replicativo del bacteriófago  $\phi 29$  es un sistema robusto y muy eficiente sobre el que existe un importante volumen de información. El estudio de las primeras etapas del proceso de replicación, central para la comprensión de la amplificación cebada por TP, es una pieza clave para el paso de la investigación básica a la aplicada. Por esa razón, se ha realizado un estudio sobre la eficiencia de utilización de los orígenes de replicación en base a cambios en su secuencia con un doble objetivo: por un lado el de obtener información sobre el proceso de reconocimiento de los orígenes de replicación por el complejo TP/DNA polimerasa y, por otro, el de aumentar la eficiencia de los primeros pasos de la replicación y optimización de la amplificación basada en TP.

Para llevar a cabo el estudio, se simplificaron los orígenes de replicación del genoma de  $\phi 29$  mediante el diseño de orígenes artificiales romos de dsDNA contruidos hibridando oligonucleótidos de 29 pb, o 68 pb en algunos casos, para producir diferentes secuencias y estructuras terminales (ver *Materiales y Métodos*, **Tabla 2**). Puesto que la DNA polimerasa tiene una potente actividad exonucleolítica 3'-5' se optó por proteger los dos últimos enlaces en 3' de cada cadena con enlaces fosforotioato con el fin de evitar su degradación (Skerra, 1992). Por otro lado, todos los orígenes artificiales cuentan con un grupo fosfato en posición 5' como consecuencia de la información obtenida en el capítulo anterior sobre su importancia. Las condiciones experimentales para los orígenes artificiales se pusieron a punto previamente y son las indicadas en *Materiales y Métodos*. Se obtuvo por síntesis una escalera de peso molecular para asegurarnos de que el tamaño unidad de replicación correspondía a los 29 pb unidos a una TP (**Figura 15**).



## 1. Efecto de la sustitución y desapareamiento del primer triplete en la eficiencia de utilización del origen de replicación

A pesar de que el proceso de iniciación de la replicación ha sido extensamente estudiado, hay muchas secuencias y estructuras no estudiadas hasta el momento. Como se ha descrito anteriormente, el triplete terminal  $3' \text{TTT} 5'$  es necesario para el óptimo funcionamiento del mecanismo de *sliding-back* que evita la pérdida de información en dicho proceso. Por esa razón, se mutaron estas tres primeras posiciones como bloques homooligoméricos de tres nucleótidos en ambas cadenas, la molde y la desplazada, y en todas las combinaciones posibles, manteniendo el resto de la secuencia igual. Esto permite no solo analizar los efectos de secuencia, sino también del desapareamiento en aquellas construcciones con extremos no complementarios. Las construcciones se analizaron en experimentos de replicación en los que se utilizó  $[\alpha^{32}\text{P}]\text{dATP}$  para minimizar el error en la cuantificación de las bandas correspondientes al producto de replicación (**Figura 16**).



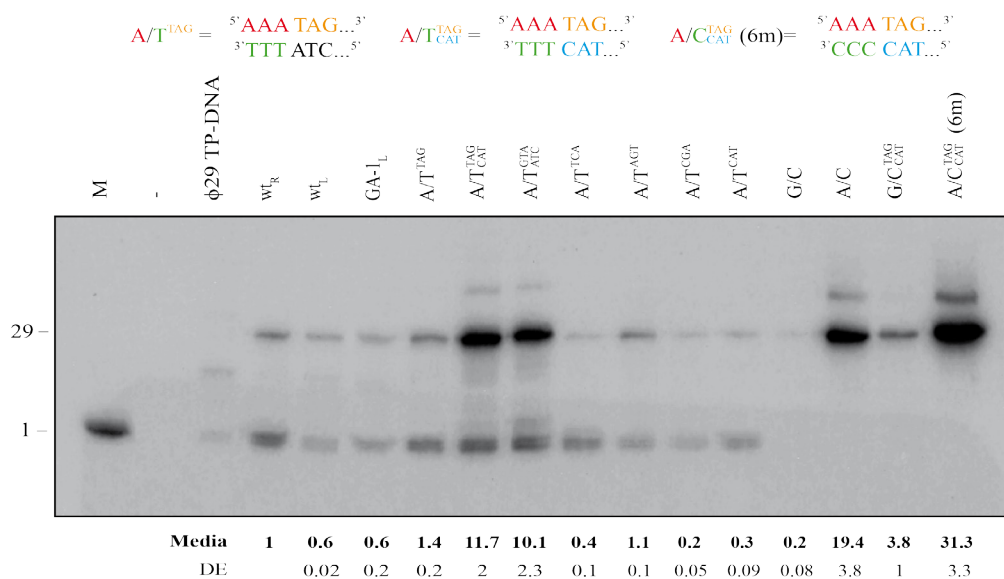
**Figura 16. Efecto de la sustitución y desapareamiento del primer triplete en la eficiencia de utilización del origen de replicación.** Diferentes oligonucleótidos de cadena sencilla se combinaron para conformar orígenes artificiales de doble banda con diferentes secuencias y desapareamientos en el primer triplete. La reacción de replicación se realizó y procesó como se describe en *Materiales y Métodos* empleando  $[\alpha^{32}\text{P}]\text{dATP}$  como nucleótido marcado. Se indican las posiciones correspondientes con la iniciación TP-dAMP [+1,+2] y longitud completa [+29] (TP-29 nucleótidos ssDNA). Carril M, marcador de iniciación (empleando el TP-DNA de  $\phi 29$  con dATP como único nucleótido); carril  $\phi 29$  TP-DNA, control de replicación (empleando TP-DNA de  $\phi 29$  como molde y todos los nucleótidos), se observan las bandas de iniciación [+1, +2], de elongación parcial [+15, +16] y de longitud completa [no mostrada]). Los diferentes pares de oligonucleótidos hibridados para generar los orígenes artificiales se nombran como sigue (ver esquema superior): la primera letra corresponde a los tres primeros nucleótidos de la hebra desplazada y la segunda para los tres primeros de la hebra molde (por ejemplo, A/T supone un origen compuesto por la secuencia pareja  $5' \text{AAA} 3' / 3' \text{TTT} 5'$  en el extremo del origen artificial). El resto de la secuencia es la secuencia *wild-type* del extremo derecho del genoma de  $\phi 29$  de la posición 4-29. El origen A/T es el origen *wild-type*. En la parte inferior se muestran los valores medios de la cuantificación de las intensidades de las bandas correspondientes a la longitud completa de replicación estandarizados con respecto al de la *wild-type* (wt=1) junto a la desviación estándar. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

En todos los casos los valores de actividad se estandarizaron incorporando a los datos obtenidos el número de As que cada uno de los orígenes incorporaba en su replicación total. Los resultados (**Figura 16**) mostraron una eficiencia de replicación muy baja para aquellas construcciones con purinas en el triplete de la hebra molde (N/A o N/G), intermedia para los moldes *wild-type* (N/T) y la más alta para los moldes con C (N/C), a excepción del par GGG/CCC (G/C). Con respecto a la hebra desplazada, no hay diferencias significativas en las construcciones que tienen la secuencia *wild-type* en la hebra molde, independientemente de que estén apareadas o no. Sin embargo, en las construcciones con C en el triplete molde se observan grandes diferencias dependiendo de la secuencia en la hebra desplazada. La mayor actividad se obtiene con el triplete AAA en la hebra desplazada (carril A/C, 18 veces la actividad *wild-type*), más baja pero aún 7 veces superior a la *wild-type* en construcciones con TTT o CCC en el triplete de la hebra desplazada (carriles T/C y C/C, 7.8 y 6.7 veces la actividad *wild-type*, respectivamente) y una muy baja para el par G/C (0.5 veces la actividad *wild-type*). En el contexto del molde con C, esto indica un efecto negativo para el par GGG/CCC mientras que uno positivo para el par AAA/CCC, posiblemente debido a que la unión del par GC es más fuerte que la del AT y, por tanto, más difícil de separar para el acceso de la maquinaria replicativa. A pesar de que el nucleótido marcado empleado es dATP y no dGTP, aparecen dos bandas de iniciación débiles en el carril A/C, que se consideran errores de incorporación dado el alto nivel de producción de TP-DNA que se obtiene con dicho origen. De estos resultados se concluye que la identidad de secuencia del primer triplete de la hebra molde es importante y que, en el caso particular de los moldes con C, se observa un efecto desfavorable en el apareamiento y una clara preferencia por A en el triplete de la hebra desplazada.

## 2. Efecto de la sustitución y desapareamiento de los nucleótidos 4-6 en la eficiencia de utilización del origen de replicación

La presencia de ITRs en el genoma de  $\phi 29$  hace que los extremos izquierdo y derecho del genoma compartan la misma secuencia en los primeros 6 pb. Sin embargo, a partir de la posición 7 existe una gran variabilidad. Por esta razón, se compararon ambas secuencias mediante el diseño de orígenes artificiales de 29-mer correspondientes a las secuencias de los extremos derecho e izquierdo, resultando la eficiencia de la primera ligeramente superior a la de la segunda (**Figura 17**, carriles wt<sub>R</sub> y wt<sub>L</sub>). Esto indica que la secuencia nucleotídica a partir de la posición 6 tiene un efecto en la utilización del origen de replicación como ya se había observado previamente (Gutiérrez *et al.*, 1986).

Mientras que la repetición terminal de tres nucleótidos suponía una ventaja considerable a la hora de realizar el análisis mutacional, la complejidad de la secuencia a partir de la posición 3 requiere la



**Figura 17. Efecto de la sustitución y desapareamiento de los nucleótidos 4-6 en la eficiencia de utilización del origen de replicación.** Diferentes oligonucleótidos de cadena sencilla se combinaron para conformar orígenes artificiales de doble banda con las diferentes secuencias y desapareamientos. Se indican posiciones correspondientes a la iniciación TP-dAMP [+1] y longitud completa [+29] (TP-29 nucleótidos ssDNA). Carril wt<sub>R</sub>, origen artificial con la secuencia del extremo derecho del genoma de ϕ29 (29 pb); carril wt<sub>L</sub>, origen artificial con la secuencia del extremo izquierdo del genoma de ϕ29 (29 pb); carril GA-1<sub>L</sub>, origen artificial compuesto por los 12 primeros pb del genoma de GA-1 y el resto del extremo derecho de ϕ29. Para el resto de carriles, como en la **Figura 16**, las dos letras mayúsculas corresponden a los nucleótidos del triplete de la hebra desplazada y molde, respectivamente. Las letras en el superíndice corresponden a la secuencia en las posiciones 4-6 de la banda desplazada y sus pares en la molde si son diferentes a la secuencia *wild-type*. En el caso de que se hayan escogido dos secuencias que no hibridan entre sí, la secuencia en las posiciones 4-6 de la hebra molde se muestran con el subíndice. En la parte inferior se muestran los valores medios de la cuantificación de las intensidades de las bandas correspondientes a la longitud completa de replicación estandarizados con respecto al de la *wild-type* (wt=1) junto a la desviación estándar. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

adopción de un compromiso en su estudio dado lo grande que es el número de secuencias posibles a analizar. Por esta razón, el análisis de las posiciones 4-6 se llevó a cabo mediante una comparación de secuencia con la del bacteriófago GA-1, que pertenece a la misma familia y género que ϕ29. Como se ha mencionado en la *Introducción*, los orígenes mínimos de replicación están comprendidos en los primeros 12 pb. Si se compara la secuencia de ϕ29 con la de GA-1 en estas primeras 12 posiciones, si bien el triplete iniciador es el mismo, el extremo derecho presenta una secuencia completamente diferente mientras que en el izquierdo hay 5 cambios. Debido a ello, se analizó el efecto resultante de sustituir los primeros 12 pb de la secuencia del extremo derecho de ϕ29 por los del extremo izquierdo de GA-1, manteniendo el resto de la secuencia *wild-type* (carril GA-1<sub>L</sub>) y resultando una actividad similar a la del



extremo izquierdo de  $\phi 29$ . De forma más específica, se sustituyeron las posiciones 4-6 de  $\phi 29$  por las de GA-1 (carril A/T<sup>TAG</sup>), resultando una actividad similar a la obtenida con el extremo derecho de  $\phi 29$ . Esto indica que las posiciones 4-6 de  $\phi 29$  pueden ser intercambiadas por las de GA-1 sin un efecto negativo. Esta permisividad se estudió mediante la sustitución de dichas posiciones por cuatro secuencias aleatorias y se observó que una de ellas (5'AGT<sup>3'</sup>, carril A/T<sup>AGT</sup>) alcanzaba los niveles de actividad *wild-type* mientras que las otras tres (5'TCA<sup>3'</sup>, 5'CGA<sup>3'</sup> y 5'CAT<sup>3'</sup>, carriles A/T<sup>TCA</sup>, A/T<sup>CGA</sup> y A/T<sup>CAT</sup>, respectivamente) daban actividades menores. Por lo tanto, algunos cambios están permitidos en las posiciones 4-6 de la secuencia de  $\phi 29$  pero otros disminuyen la eficiencia de replicación, siendo en este momento difícil de predecir la eficiencia de los orígenes *a priori*. Puesto que los orígenes con la secuencia de GA-1 en las posiciones 4-6 resultaron con una actividad similar a la *wild-type*, se analizó el efecto del desapareamiento resultante de la hibridación de dicha secuencia y la de  $\phi 29$ , generando las dos combinaciones posibles (AAATAG/TTT-CAT y AAAGTA/TTTATC). Utilizando ambos orígenes se obtuvieron actividades alrededor de 10 veces superiores a la *wild-type* (carriles A/T<sup>TAG</sup><sub>CAT</sub> y A/T<sup>GTA</sup><sub>ATC</sub>), siendo la primera combinación la más eficiente. Esto sugiere que el desapareamiento en las posiciones 4-6 tiene un fuerte efecto estimulador en la utilización del origen. Teniendo en cuenta los resultados del apartado anterior, donde el par A/C era el origen más activo (**Figura 17**, carril A/C), dicho desapareamiento se combinó con el de las posiciones 4-6 que había resultado con la mayor actividad en este segundo grupo de orígenes artificiales. Dicha combinación (AAATAG/CCCCAT), con un desapareamiento de las primeras 6 pb resultó con la actividad más alta obtenida hasta el momento, 30 veces superior a la actividad *wild-type* (carril A/C<sup>TAG</sup><sub>CAT</sub> o 6m). Si, en este caso, se cambia el triplete de la hebra desplazada por G, formando un par GGG/CCC en el extremo, la actividad se reduce drásticamente hasta 3.8 veces la actividad *wild-type* (carril G/C<sup>TAG</sup><sub>CAT</sub>). Estos resultados sugieren que: (i) desapareamientos de 6 posiciones son permitidos y generan orígenes más activos; (ii) algunos desapareamientos pueden ser combinados de forma sinérgica; y (iii) el efecto positivo observado en el desapareamiento del extremo podría ser producto de la superación de alguna barrera energética a ese nivel.

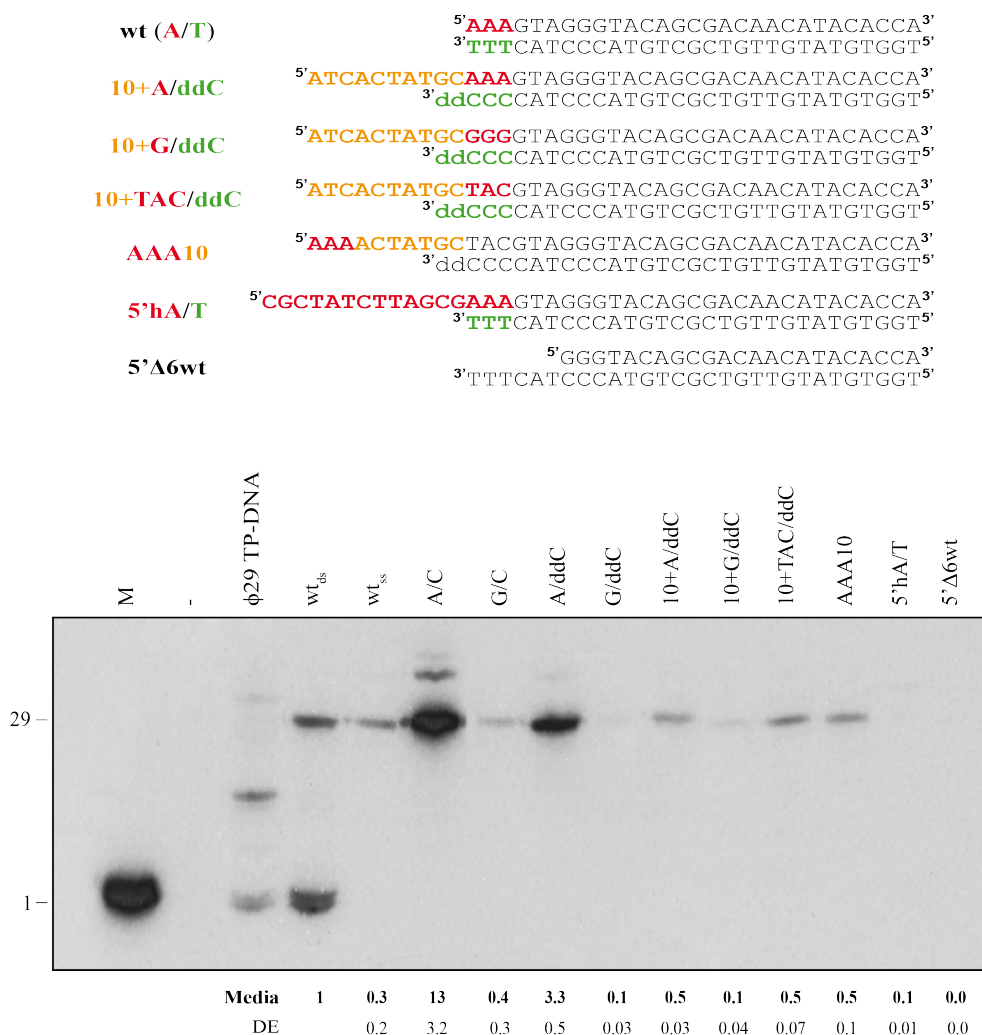
**Tabla 3. Efecto de la secuencia en orígenes artificiales con 6 desapareamientos en el extremo.** Los diferentes orígenes artificiales se construyeron mediante la hibridación de los correspondientes oligonucleótidos de banda simple. Sobre los orígenes se llevó a cabo un experimento de replicación como se describe en *Materiales y Métodos*. Las secuencias de la primera columna corresponden a los primeros 6 nucleótidos de las hebras desplazada y molde, respectivamente, siendo el resto de la secuencia igual al del origen wt. La segunda columna (Media) corresponde a las actividades de los diferentes orígenes como el porcentaje estandarizado con respecto a la construcción 6m (6m=100%) para facilitar su comparación. En la tercera columna se muestran las DE correspondientes a los porcentajes de la segunda columna. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

		Media (%)	DE
(wt)	AAAGTA TTTCAT	3	1.2
(G/C)	GGGGTA CCCCAT	0.7	0.1
(A/C)	AAAGTA CCCCAT	49.2	12
(6m)	AAATAG CCCCAT	100	
	AAACAT CCCCAT	109.5	17.8
	AAACGA CCCATC	66.3	10.5
	AAAGTA CCCATC	58.1	4.1
	AAAAGT CCCATC	57.1	7.5
	AAATAG CCCCGT	49.1	3.3
	AAACGA CCCCGTA	43.8	6.3
	AAAGTA CCCCGTA	43.1	3.9
	AAATCA CCCCGTA	27.6	6
	TTTTAG CCCCAT	64.9	14.1
	CTCACG CCCCAT	71.1	12.3

Puesto que la mayor actividad de replicación se obtuvo con un origen con las 6 primeras posiciones desapareadas, se exploró dicho desapareamiento mediante la combinación de algunas secuencias al azar que mantuvieran esa apertura. Los resultados se observan en la **Tabla 3**. Considerando la actividad del origen 6m del 100%, las actividades variaron desde un 27% a 109% (puesto que el origen 6m es 30 veces más eficiente que el wt, la menos eficiente de las construcciones de 6 desapareamientos analizada es todavía 10 veces más eficiente que el wt). A pesar de la preferencia observada anteriormente, algunas construcciones sin un triplete AAA en la hebra desplazada alcanzaban un 65% (TTT) o un 71% (secuencia aleatoria 5'CTCAGC3'). Estos resultados apuntan a que, si bien el desapareamiento mejora fuertemente la utilización del origen, la secuencia de las hebras individuales también tiene una gran influencia en la actividad del origen. Incluso se obtienen orígenes muy activos que no comparten la secuencia en ninguna de las posiciones con la *wild-type*. De los datos obtenidos es muy difícil predecir, en orígenes que difieran mucho de los empleados en este estudio, qué secuencias van a constituir orígenes más o menos activos.

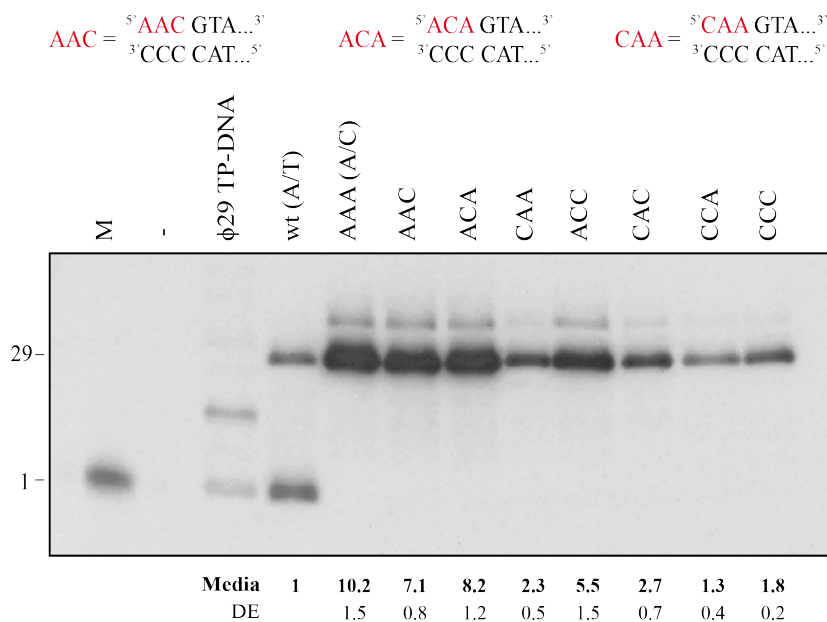
### 3. Efecto de protuberancias en el extremo 5' en la eficiencia de utilización del origen de replicación

Trabajos anteriores demostraron que la maquinaria replicativa de  $\phi$ 29 era capaz de utilizar como molde un DNA de banda simple (Blanco *et al.*, 1989). Además, como se ha mostrado en el capítulo anterior, gracias a la actividad exonucleolítica 3'-5' intrínseca de la DNA polimerasa, esta puede eliminar extensiones de ssDNA en 3' para generar extremos romos tipo *wild-type* susceptibles de ser amplificados (Mencia *et al.*, 2011). El efecto positivo observado en moldes con las primeras posiciones desapareadas podría deberse a que un incremento en el número de interacciones del complejo DNA polimerasa/TP con la hebra desplazada podría estabilizar el complejo de iniciación produciendo un aumento en la eficiencia de utilización. Por esa razón se analizó el efecto de extensiones de banda simple en los extremos 5' de distintos orígenes con secuencia y estructuras terminales representativas que habían sido anteriormente analizadas. Se diseñaron diferentes orígenes con extensiones de 10 nucleótidos en los extremos 5' y se escogió una longitud de 10 nucleótidos para proporcionar suficiente extensión de ssDNA como para cubrir, probablemente, una buena parte del complejo DNA polimerasa/TP. Puesto que el extremo recesivo 3' resultante podría ser un buen molde para la elongación por la DNA polimerasa libre, afectando al resultado del experimento, se diseñaron oligonucleótidos con dideoxi-CMP (ddC) en la última posición de la hebra molde que impidieran su extensión. Como control de un posible efecto del ddC, se probaron las construcciones A/C y G/C con y sin ddC. Como se observa en la **Figura 18**, los orígenes A/C y G/C con ddC (carriles A/ddC y G/ddC, respectivamente) resultaron con un 25% de la actividad total de sus respectivos



**Figura 18. Efecto de extensiones de banda simple en los extremos 5' en la utilización de los orígenes de replicación.** Se generaron diferentes estructuras terminales mediante la hibridación de los correspondientes oligonucleótidos de banda simple. La reacción de replicación se realizó y procesó como se describe en *Materiales y Métodos* empleando [ $\alpha^{32}$ P]dATP como nucleótido marcado. Carril wt<sub>ds</sub>, origen con la secuencia del extremo derecho del genoma de  $\phi$ 29 (29 pb); carril wt<sub>ss</sub>, origen con la secuencia de la hebra molde del extremo derecho del genoma de  $\phi$ 29; los carriles A/ddC y G/ddC siguen la misma nomenclatura que en la **Figura 16**, salvo que el primer nucleótido en el extremo 3' de la hebra molde es un dideoxiCMP; carril 10+A/ddC, extensión de banda simple de 10 nucleótidos y secuencia 5'ATCACTATGC3' en el extremo 5' de un origen A/ddC; carril 10+G/ddC, la misma extensión en un origen G/ddC; carril 10+TAC/ddC, la misma extensión en un origen con secuencia TAC/ddCCC en el extremo; carril AAA10, origen con una extensión de banda simple de 10 nucleótidos y secuencia 5'AAACTATGC3' en el extremo 5' de un origen TAC/ddC; carril 5'hA/T, extensión en banda simple de 13 nucleótidos y secuencia 5'CGCTATCTTAGCG3' en el extremo 5' que forma una horquilla autocomplementaria de 5 pb en un origen wt; carril 5'Δ6wt, origen wt con una delección de 6 nucleótidos en la hebra desplazada. Se muestran los valores medios de la cuantificación de las intensidades de las bandas correspondientes con la longitud completa de replicación estandarizados con respecto al de la *wild-type* (wt=1) junto a la desviación estándar. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

orígenes homólogos sin ddC. Aunque la razón de esto es desconocida, una explicación podría ser que el grupo hidroxilo en 3' pudiera ser importante en la interacción con la TP en su primera asociación con el DNA. Por otra parte, se observó que la actividad obtenida con el origen *wild-type* (carril wt<sub>ds</sub>) era 3 veces superior al origen *wild-type* en banda simple (carril wt<sub>ss</sub>), lo que pone de manifiesto de nuevo la importancia de la hebra desplazada en los eventos tempranos de la replicación. Para el diseño de los orígenes con extremos 5' protuberantes se tomó una secuencia de 10 nucleótidos al azar (5' ATCACTATGC3') y se añadió a un molde AAA/ddC (carril 10+A/ddC), a un molde GGG/ddC (carril 10+G/ddC) y a un molde TAC/ddC (10+TAC/ddC). Puesto que se había observado una preferencia por A en el extremo 5', también se diseñó una extensión de 10 nucleótidos con un triplete de A en el extremo 5' (5' AAACTATGC3') sobre un molde TAC/ddC (carril AAA10). Con todos los orígenes con extensiones en 5' se obtuvo una actividad menor a la de los orígenes A/ddC y *wild-type*. Como se esperaba, el origen que presentaba el segmento de doble cadena completamente hibridado mostró una actividad aún menor que los otros (carril 10+G/ddC). Por otra parte, un origen *wild-type* con una extensión en 5' con la secuencia 5' CGCTATCTTAGC3' que autohibrida en forma de horquilla (carril 5'hA/T) y un origen *wild-type* al que se le han eliminado los 6 primeros nucleótidos de la hebra desplazada (carril 5'Δ6wt) fueron analizados, sin mostrar actividad. En el caso de la construcción 5'hA/T (ver **Figura 18**), probablemente la horquilla formada constituya un impedimento estérico que impida el acceso del heterodímero TP/DNA polimerasa al par A/T. En el caso de la construcción 5'Δ6wt (ver **Figura 18**), tanto el triplete TTT del extremo de la banda simple como el par G/C podrían ser empleados para iniciar la replicación. Sin embargo, ninguno de los dos es un molde eficiente como se ha visto en resultados previos. Posiblemente, el molde en doble banda sea preferido por el heterodímero DNA polimerasa/TP para la iniciación de la replicación, pero la extensión en banda simple del extremo 3' hace que la hebra molde no pueda encajar en el bolsillo catalítico configurado por el heterodímero. Por otra parte, la extensión en 3' es un buen sustrato para la actividad exonucleasa de la DNA polimerasa, por lo que el enzima podría estar bloqueada tratando de degradar improductivamente el extremo 3' durante el tiempo de reacción aun estando protegido por el fosforotioato. Estos resultados muestran de nuevo la importancia de la hebra desplazada en la eficiencia de uso del origen de replicación y el hecho de que cuando esta es prolongada artificialmente para favorecer el número de interacciones con el heterodímero, no se produce ninguna estimulación, al contrario de lo que sucede cuando el dsDNA es abierto de forma artificial mediante la utilización de oligonucleótidos con extremos no complementarios.



**Figura 19. Caracterización de la preferencia del heterodímero DNA polimerasa/TP por AAA en el extremo 5' de la hebra desplazada.** Los orígenes se generaron mediante la hibridación de los correspondientes oligonucleótidos de banda simple, siendo la construcción A/C la base de todas las modificaciones. La reacción de replicación se realizó y procesó como se describe en *Materiales y Métodos* empleando [ $\alpha^{32}\text{P}$ ]dATP como nucleótido marcado. La nomenclatura NNN indica la secuencia de los tres primeros nucleótidos en el extremo 5' de la hebra desplazada siendo en todos los casos  $3'\text{CCC}5'$  en la hebra molde. Se muestran los valores medios de la cuantificación de las intensidades de las bandas correspondientes con la longitud completa de replicación estandarizados con respecto al de la *wild-type* (wt=1) junto a la desviación estándar. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

#### 4. Preferencia del heterodímero TP/DNA polimerasa por la secuencia AAA en el extremo 5' de la banda desplazada

A lo largo del estudio, se han obtenido evidencias de que existía una preferencia por el triplete AAA en el extremo 5' de la hebra desplazada. Además, las mayores actividades observadas (6m; AAACAT/CCCCAT) se obtenían en orígenes con el desapareamiento AAA/CCC en el extremo. Por esa razón, se estudió la preferencia en cada una de las tres posiciones de forma separada a partir de la construcción original A/C (A/C corresponde a AAA en la **Figura 19**), puesto que es la que mostraba un efecto más claro de las construcciones con cambios en la secuencia de la hebra desplazada respecto a la *wild-type*. Se sustituyeron todas las A en la banda desplazada de forma individual o en combinaciones por C. Los resultados (**Figura 19**) mostraron que la sustitución de la primera A suponía una caída en la actividad de 5 veces (compárese carriles AAA con CAA). La sustitución de la segunda A reducía la actividad replicativa de un

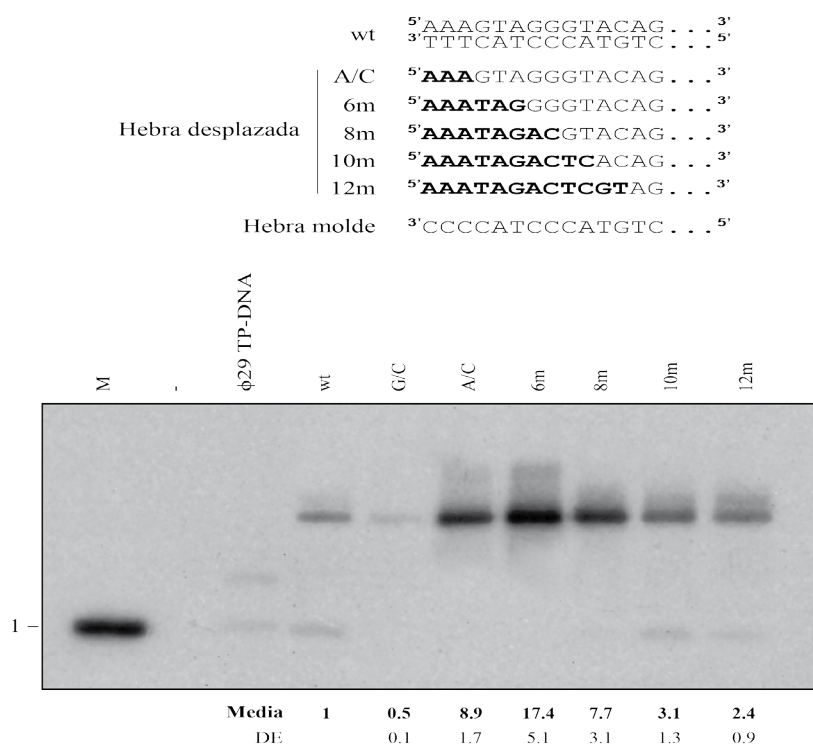
20% a un 50% (compárese carriles AAA con ACA, y CAC con CCC), es decir, el efecto era menor que el observado en la primera posición. Sin embargo, en el caso de las sustituciones de la tercera posición, en algunos orígenes se observaba una pequeña reducción en la actividad (compárese carriles AAA con AAC, y ACA con ACC), mientras que en otros casos se observaba un ligero aumento (compárese carriles CAA con CAC, y CCA con CCC). Si se comparan los orígenes con una sola A en el triplete 5' (ACC, CAC, CCA), los resultados indican que se puede establecer una jerarquía de importancia de las tres posiciones: la primera A es más importante que la segunda, y esta que la tercera, pero que cada una de las posiciones del triplete en 5' parece tener su efecto.

### 5. Efecto de la apertura de la hélice en la eficiencia de utilización del origen de replicación

Puesto que el origen con el que se obtuvo la mayor actividad constaba de 6 posiciones desapareadas (6m), un mayor número de desapareamientos podría suponer aún un mayor nivel de estimulación. Para dicho análisis, se diseñaron orígenes con un total de 8, 10 y 12 desapareamientos. Para minimizar el posible efecto de secuencia, se diseñaron orígenes que, a pesar de la apertura de la hélice, compartían la misma secuencia en la hebra molde. De esta forma, el origen 8m tenía un total de 8 desapareamientos, de los que los 6 primeros nucleótidos eran iguales al origen 6m y dos adicionales, así como el origen 10m compartía la secuencia de los 8 desapareamientos del origen 8m y el 12 los 10 del origen 10m.

Los desapareamientos en la doble banda podrían suponer un problema en la estabilidad del origen por tener una porción considerable del mismo desapareada. Por esa razón, en este experimento, se emplearon oligonucleótidos de 68 nucleótidos en la construcción de los orígenes artificiales, ya que el efecto de los desapareamientos con esta longitud y a la temperatura empleada es mínimo. Puesto que se había aumentado la longitud de los orígenes considerablemente, se llevó a cabo un experimento control donde se comparaban las actividades de replicación de moldes 29mer y 68mer (**Figura S1**). Se observó que, mientras el origen wt presentaba una actividad similar en sus versiones 29-mer y 68-mer, la actividad de las versiones 68-mer con respecto a las 29-mer de los orígenes A/C y 6m, caían 1.4 y 2.9 veces, respectivamente. Los resultados (**Figura 20**) muestran que el desapareamiento óptimo para la eficiencia de replicación del origen es de 6 nucleótidos. Esta actividad es 2 veces menor que la que se obtenía con el origen 6m 29mer, aunque se desconoce la razón de este efecto. A medida que la hélice se va abriendo, la actividad disminuye progresivamente hasta alcanzar 2.4 veces la actividad *wild-type* con la construcción 12m.





**Figura 20. Número óptimo de desapareamientos en el extremo para la utilización del origen de replicación.** Las construcciones se analizaron en un experimento de replicación como se describe en *Materiales y Métodos*. En este caso, los orígenes se diseñaron con una longitud de 68 pb para eliminar posibles problemas en la estabilidad del origen. La nomenclatura para wt, G/C, A/C y 6m es la empleada en anteriores figuras. El número seguido de 'm' indica el número de posiciones desapareadas en el origen. Los desapareamientos se diseñaron de forma que se fueron sustituyendo progresivamente las posiciones 7-12 de la hebra desplazada (secuencia 5'AAATAG<sup>6</sup>GGGTAC<sup>3</sup>) por la secuencia no complementaria 5'ACTCGT<sup>3</sup>. De esa forma, la secuencia de los desapareamientos de la hebra desplazada es, para el origen 8m 5'AAATAGAC<sup>3</sup>, para el origen 10m 5'AAATAGACTC<sup>3</sup>, y para el origen 12m 5'AATAGACTCGT<sup>3</sup>. Se muestran los valores medios de la cuantificación de las intensidades de las bandas correspondientes con la longitud completa de replicación estandarizados con respecto al de la *wild-type* (wt 68-mer=1) junto a la desviación estándar. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

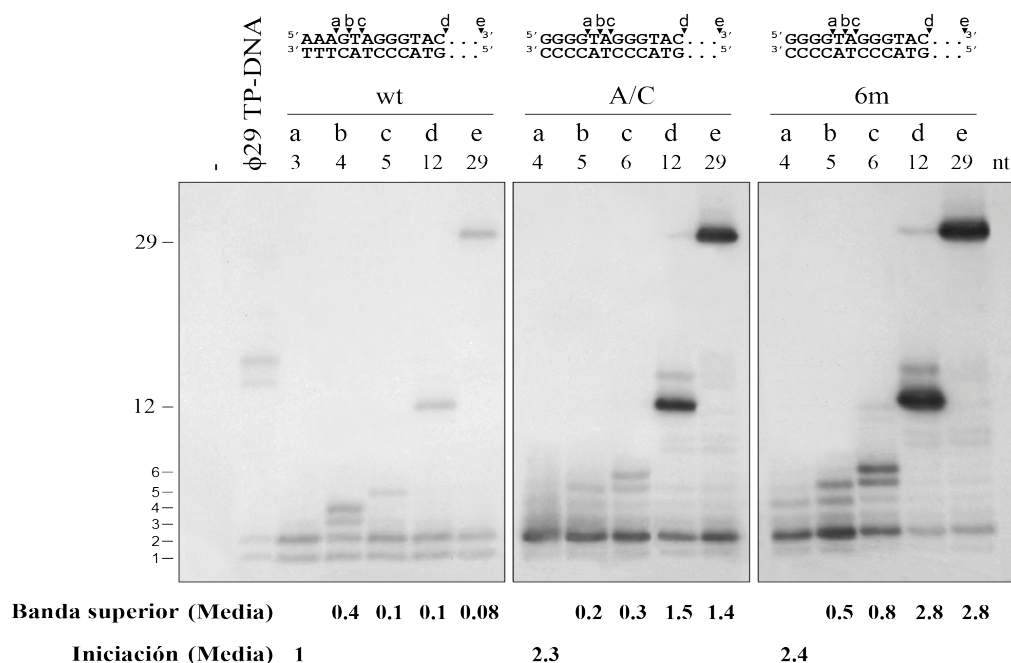
## 6. Pasos afectados en la replicación por los orígenes de replicación desapareados

Como se ha explicado en la *Introducción*, la replicación del fago φ29 comienza por una etapa de iniciación seguida por una etapa de transición hasta que la hebra de nueva síntesis alcanza una longitud de unos 9 nucleótidos. A partir de la incorporación del nucleótido 10, se producen en la DNA polimerasa ciertos cambios estructurales que conllevan a la disociación entre la DNA polimerasa y la TP, procediendo con la elongación de la hebra de nueva síntesis. Como se ha observado, con algunos de los orígenes artificiales

estudiados se obtenían actividades de replicación mayores que con el origen *wild-type*, especialmente aquellas con el par AAA/CCC en el extremo. Para analizar qué paso o pasos dentro del proceso replicativo estaban siendo facilitados en las etapas de iniciación y elongación por estos nuevos orígenes, se realizó un experimento de transición (véase *Materiales y Métodos*) con los orígenes A/C y 6m (**Figura 21**). En dicho experimento, las reacciones contenían: carriles a, el primer nucleótido incorporado (produciendo AAA o GGGG como bandas de iniciación, dependiendo del origen molde); carriles b, el primer nucleótido (A o G) más ddG (produciendo AAAG en el origen *wild-type*) o más ddT (produciendo GGGGT en los orígenes A/C y 6m); carriles c, los nucleótidos A y G más ddT (produciendo AAAGT en el origen *wild-type*) o los nucleótidos G y T más ddA (produciendo GGGGTA en los orígenes A/C y 6m); carriles d, nucleótidos A, G y T más ddC (produciendo la polimerización de 12 nucleótidos en todos los orígenes); y carriles e, los cuatro nucleótidos para la replicación total del origen. Las cuantificaciones y posteriores análisis estadísticos se realizaron, al igual que en todos los otros experimentos, teniendo en cuenta la cantidad de nucleótidos marcados radiactivamente que se habían incorporado en cada origen. Se observó que, en el caso del origen *wild-type*, las bandas de iniciación (carril a) progresaban con eficacia a la posición +4 (carril b), pero con muy poca a las posiciones +5, +12 o la replicación total (carriles c, d y e), lo que indica que la transición no es un proceso eficiente. Para los orígenes A/C y 6m, se observó un incremento en el producto de iniciación de 2.3 y 2.4 veces, respectivamente, con respecto al producido con el origen *wild-type*. Pero además, se observó un fuerte incremento en el paso a +12 (11 veces para A/C, 21 veces para 6m) o a la longitud total, especialmente en el caso del origen 6m. Los resultados sugieren que los orígenes con el par AAA/CCC en el extremo están facilitando el paso de iniciación, así como estimulando fuertemente el paso de transición.

En este experimento, las bandas que aparecen por debajo del tamaño unidad de replicación (frecuentemente +1 o +2) podrían suponer eventos de iniciación abortiva o productos de un equilibrio polimerización-exonucleólisis. En todos los casos se deduce la dificultad de la polimerasa para pasar de la iniciación a la transición. En el origen *wild-type*, la cantidad relativa de productos +12 o +29 con respecto a +1 o +2 muestra una baja eficiencia en la generación de productos de longitud completa. En los orígenes A/C y, especialmente, 6m, además de un alto nivel de iniciación, la eficiencia es mucho mejor que la observada con el origen *wild-type*. Esta mejora podría ser debida a dos posibilidades, por la reducción de la generación de productos abortivos o por el desplazamiento del equilibrio exonucleólisis-polimerización hacia la polimerización. La explicación más simple para ambas situaciones es que la etapa de transición está siendo facilitada para la DNA polimerasa, de modo que la mayor parte de las iniciaciones progresan a longitud completa, sin detenerse ni degradarse. En el caso particular del origen 6m, según se proporcionan los nucleótidos se observa un descenso en los productos de iniciación (productos +1/+2), a la vez que un aumento en la relación producto replicación:productos de iniciación, lo que es consistente con la idea de que el proceso de transición está siendo facilitado.



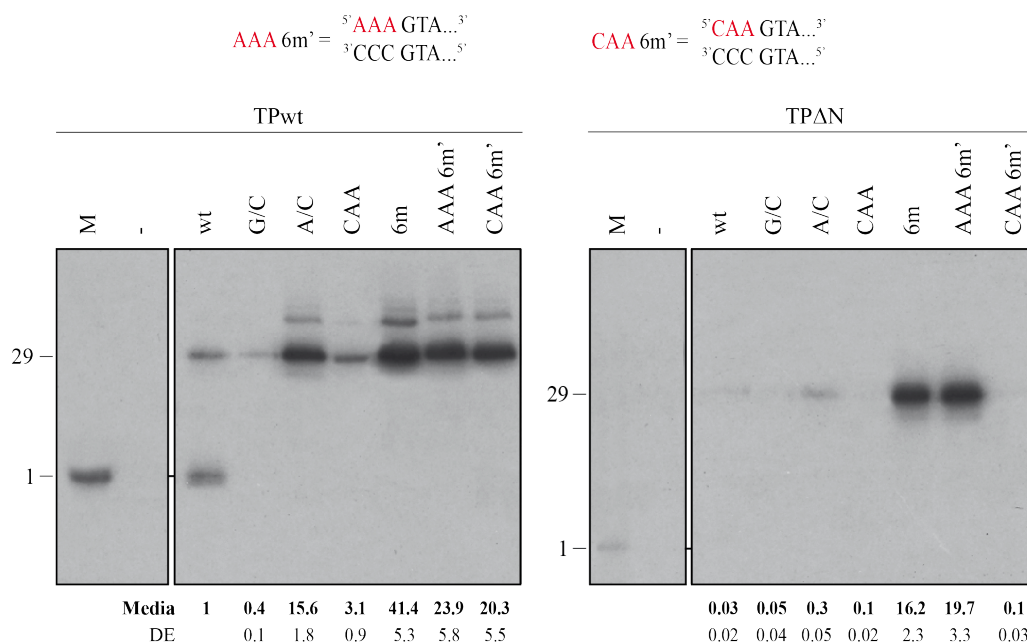


**Figura 21. Pasos replicativos afectados por los orígenes artificiales.** El experimento de transición se realizó y procesó como se describe en *Materiales y Métodos*. Los orígenes artificiales wt, A/C y 6m se analizaron en presencia de distintas combinaciones de nucleótidos que limitan el tamaño del producto sintetizado. En la parte superior, se muestra encima de cada sustrato de DNA la secuencia de la hebra molde y su complementaria con flechas que indican la longitud del producto sintetizado en cada carril, señalado con una letra. Los números del panel superior que acompañan a las letras indican el tamaño máximo del producto de DNA en cada carril. Se muestran los valores medios de la cuantificación de las intensidades de las bandas superiores de cada carril en relación a la intensidad del producto correspondiente a la iniciación del origen *wild-type* (wt=1) junto a la desviación estándar. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

## 7. Papel del dominio N-terminal de la TP de φ29 en la replicación de los orígenes artificiales

Se ha documentado que el dominio N-terminal de la TP de φ29 tiene capacidad de unión a DNA inespecífica de secuencia y que es parcialmente prescindible para la replicación *in vitro* del DNA de φ29, aunque es importante *in vivo* (Holguera *et al.*, 2015). Para entender su función en los primeros pasos de la replicación, se ensayó un mutante de delección de la TP al que se le había eliminado el dominio N-terminal (TPΔN) para observar su comportamiento con diferentes orígenes artificiales de doble banda de los caracterizados anteriormente. Como se ha mostrado previamente, la identidad de la

secuencia en el extremo 5', así como el desapareamiento, afecta a la eficiencia de la replicación por lo que quizá el dominio N-terminal podría estar jugando un papel importante en su reconocimiento. Para el análisis se realizaron experimentos de replicación con la TPΔN y diferentes orígenes con cambios, entre ellos el primer nucleótido del extremo 5'. Para utilizar las mismas secuencias que en la hebra desplazada de la **Figura 19** (donde se determinaba la importancia del triplete AAA de la hebra desplazada), se construyó un origen con las 6 primeras bases desapareadas y la secuencia AAAGTA/CCCGTA (AAA6m') y el mutante al que se sustituyó la primera A por una C en el extremo 5' CAAGTA/CCCGTA (CAA6m'). La secuencia de las posiciones 4-6 de estos orígenes era algo distinta a la que presentaba el origen 6m, ya que se emplearon los mismos oligonucleótidos que correspondían a la hebra desplazada de la **Figura 19** y se pusieron en un contexto de 6 desapareamientos. Como se observa en la **Figura 22**, la actividad de replicación observada con el origen *wild-type* es el utilizado (carriles wt) es mucho más reducida en el caso de la TPΔNt, aproximadamente 30 veces menos que con la TP wt. De igual forma, los orígenes A/C y CAA (ambos con 3 posiciones desapareadas) mostraron una actividad muy baja (carriles A/C y CAA) en presencia de TPΔN. Sin embargo, los orígenes con los 6 primeros pares de bases desapareados mostraron con la TPΔN una actividad del 50% la de la TP wt (carriles 6m y AAA 6m'), siendo esta actividad totalmente dependiente de la A en el extremo 5' (carril CAA 6m'). Mientras que con la TP wt la replicación no era dependiente de la primera A en 5' (carril CAA 6m'), probablemente a consecuencia del contexto de 6 desapareamientos, con el mutante TPΔN se observó una pérdida total de actividad al sustituir la A en 5'. De estos resultados podemos concluir que: (i) el dominio N-terminal es importante para la utilización de moldes apareados o con desapareamientos de 3 posiciones, probablemente por su interacción con el DNA, (ii) el dominio N-terminal es menos importante para la utilización de los orígenes con desapareamientos de 6 posiciones con A en el extremo 5' de la hebra desplazada y (iii) el dominio N-terminal no se requiere para la preferencia por la A en 5' observada anteriormente para el heterodímero TP/DNA polimerasa. De hecho, la preferencia por A en 5' se ve acentuada con la proteína TPΔN. Estos resultados son acordes con la evidencia que sugiere que el dominio N-terminal se requiere para una unión eficiente a la doble cadena de DNA, sin especificidad de secuencia (Muñoz-Espín *et al.*, 2010) mientras que este es prescindible en orígenes desapareados. De esta forma, los resultados también implican que los otros dominios de la TP (Intermedio, C-terminal) y/o regiones de la DNA polimerasa de φ29 serían los responsables de la preferencia por la A en 5' e interaccionarían con los extremos desapareados.



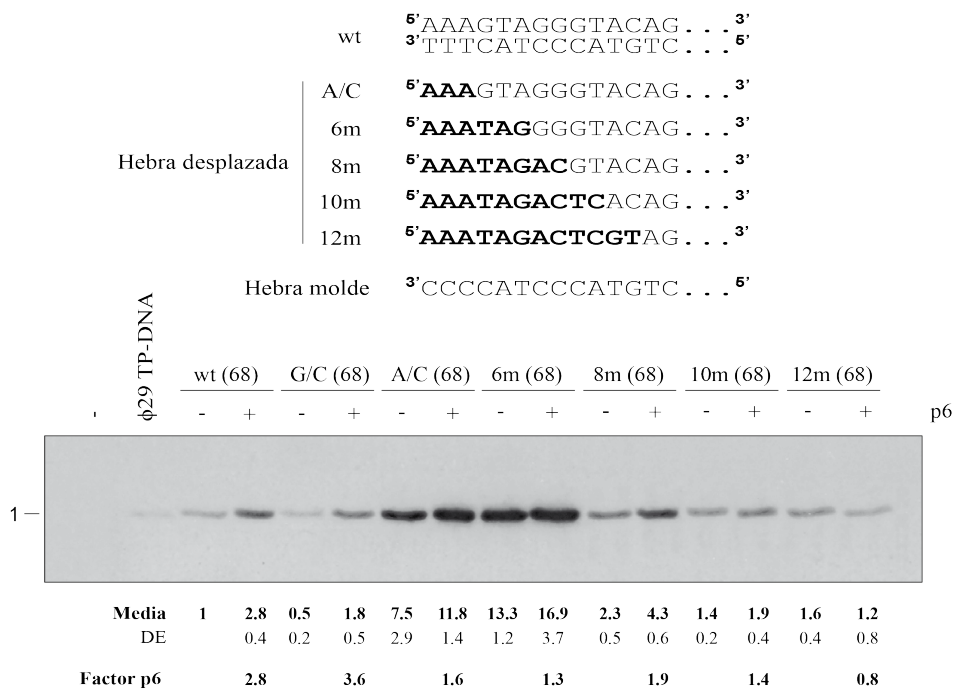
**Figura 22. Papel del dominio N-terminal de la TP de  $\phi$ 29 en la replicación de los orígenes artificiales.**

Las construcciones se analizaron en un experimento de replicación como se describe en *Materiales y Métodos*. Las reacciones se llevaron a cabo con 150 nM tanto de la TP wild-type (TP wt) como del mutante de delección (TPΔN). La nomenclatura para wt, G/C, A/C, CAA y 6m es la empleada en anteriores figuras. Carril M, marcador de iniciación para cada TP (empleando el TP-DNA de  $\phi$ 29 con dATP como único nucleótido); carril AAA6m', origen con secuencia AAAGTA/CCCGTA en el extremo; carril CAA6m', origen con secuencia CAAGTA/CCCGTA en el extremo. Se muestran los valores medios de la cuantificación de las intensidades de las bandas correspondientes a la longitud completa de replicación estandarizados con los del origen *wild-type* y la TP wt (wt=1) junto a la desviación estándar. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

## 8. Efecto de la proteína de unión a banda doble p6 en los orígenes de replicación artificiales

Como se ha comentado previamente en la *Introducción*, en la estimulación de la replicación por p6 han sido propuestos dos mecanismos de acción no excluyentes para esta proteína: (I) la apertura de la doble hélice de DNA en los extremos del genoma del fago y (II) el reclutamiento de la maquinaria replicativa, constituida por el heterodímero TP/DNA polimerasa, al origen de replicación mediante interacción directa proteína-proteína. Puesto que en el análisis sobre los orígenes de replicación se observó un efecto positivo en la actividad replicativa de los extremos desapareados y que dicha apertura se propone como uno de los

mecanismos de estimulación por p6, se analizó la actividad de dichas construcciones en presencia y ausencia de p6. El efecto de p6 es más evidente al nivel de la iniciación de la replicación y a una relativamente baja concentración de nucleótido (0.1  $\mu$ M), por lo que el análisis se llevó a cabo mediante experimentos de iniciación (véase *Materiales y Métodos*). Por otra parte, los sitios de alta afinidad para la nucleación de p6 están localizados entre las posiciones 35 y 58 (Serrano *et al.*, 1989), por lo que se emplearon orígenes artificiales de 68 pb con las construcciones y secuencias más importantes derivadas de este estudio.



**Figura 23. Efecto de la proteína p6 en la replicación de los orígenes artificiales.** Las construcciones se analizaron en un experimento de iniciación como se describe en *Materiales y Métodos* a una concentración de dNTPs final de 0.1  $\mu$ M. En este caso, los orígenes se diseñaron con una longitud de 68 pb, que comprende el sitio de nucleación de la proteína p6. Se analizaron las estructuras y secuencias más relevantes resultado del análisis de los orígenes 29-mer. La proteína p6 se empleó a una concentración final de 35  $\mu$ M y se indica su ausencia (-) o presencia (+) en la reacción. Se muestran los valores medios y desviación estándar de la cuantificación de las intensidades de las bandas correspondientes al complejo de iniciación con respecto al de la *wild-type* en ausencia de p6 (wt -p6=1). También se muestra el factor de estimulación por p6 (factor p6) mediante la relación entre los valores +p6/-p6 para cada origen. Los valores presentados son el resultado de la media de, al menos, tres experimentos independientes.

En la **Figura 23** se observan factores de estimulación por p6 (factor p6) de 2.8 para el origen wt, 3.6 para G/C y 1.6, 1.3, 1.9, 1.4 y 0.8 para los orígenes A/C, 6m, 8m, 10m y 12m, respectivamente. Mientras que en los orígenes apareados como wt y G/C la estimulación es notable, en los orígenes con extremos despareados la estimulación es baja. Este bajo nivel de estimulación por p6 podría deberse, al menos en

parte, a que el desapareamiento de estas construcciones está actuando en algún paso favorecido por p6. Una explicación alternativa podría ser la de que el desapareamiento disipa la fuerza de torsión generada por el complejo nucleoproteico p6-DNA afectando a su función. En los orígenes apareados, parece haber una mayor contribución por parte de p6 cuanto mayor es la fuerza del enlace en el triplete del extremo (wt respecto a G/C). A excepción del origen 12m, en todos los demás orígenes hay un cierto nivel de estimulación remanente por p6 que podría atribuirse a un mecanismo de acción adicional como una interacción directa proteína-proteína con el complejo TP/DNA polimerasa y su reclutamiento al origen de replicación como se ha sugerido previamente (Freire *et al.*, 1996; Pérez-Arnáiz *et al.*, 2007).



## Capítulo 3

**Sustitución del dominio N-terminal de la TP de  $\phi 29$  por el dominio de unión a DNA del factor de transcripción GAL4**

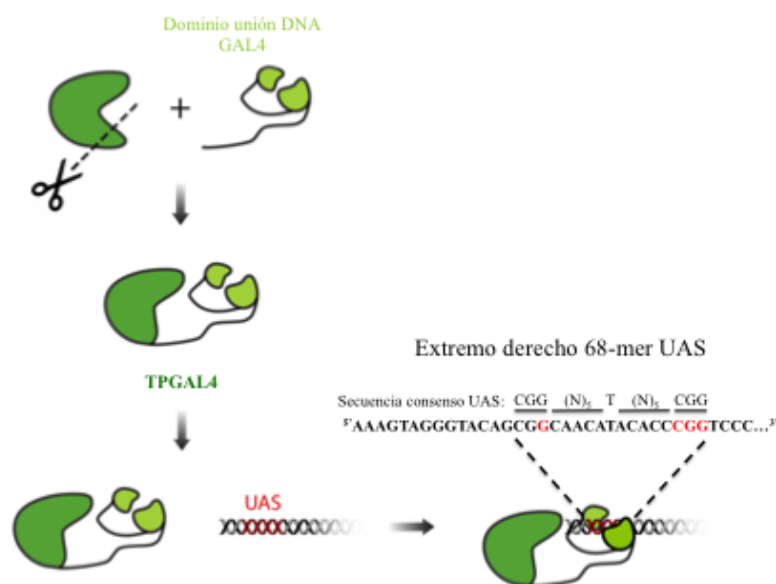




El análisis de la eficiencia de replicación de los orígenes artificiales de 29 pb y 68 pb supuso el hallazgo de una construcción que superaba en 30 veces los valores obtenidos con la secuencia original. Sin embargo, la secuencia y las estructuras terminales no son las únicas variables en la utilización de los extremos del genoma sino que, como se ha mencionado en la *Introducción*, una compleja red de interacciones proteína-proteína y proteína-DNA tienen lugar en el proceso replicativo. Aunque la estructura cristalográfica del dominio N-terminal de la TP del bacteriófago  $\phi 29$  no ha sido resuelta, se sabe que presenta capacidad de unión a DNA de forma inespecífica de secuencia y que esto es importante para su actividad (Zaballos y Salas, 1989). Para aumentar aún más la utilización de los orígenes de replicación se decidió reclutar artificialmente la TP al origen de replicación. Para ello se sustituyó el dominio N-terminal de la TP de  $\phi 29$  por el dominio de unión a DNA del factor de transcripción GAL4, que se une específicamente a la secuencia UAS (*Upstream Activation Sequence*). De esta forma se podría obtener una amplificación mejorada basada en el reclutamiento de la maquinaria de replicación mediado por el reconocimiento específico de GAL4 hacia el origen de replicación.

## 1. Diseño de la proteína TP GAL4

La **Figura 24** muestra el diseño y mecanismo de activación propuesto para la proteína TP GAL4. En su construcción se deleccionaron los primeros 73 aminoácidos correspondientes al dominio N-terminal de la TP de  $\phi 29$  para anular la unión inespecífica al DNA. Al fragmento resultante se le fusionó el dominio de unión a DNA del factor de transcripción GAL4 mediante un enlace flexible de 3 glicinas para permitir un plegamiento correcto. La proteína se expresó en bacterias *E.coli* BL21(DE3) en medio de autoinducción durante toda la noche a 30°C. La construcción y purificación se describen en *Materiales y Métodos*. Para el estudio de su actividad se diseñó un origen de 68 pb artificial basado en la secuencia *wild-type* del extremo derecho del genoma de  $\phi 29$  al que se le había introducido la secuencia consenso UAS reconocida por GAL4 (Traven, *et al.*, 2006) (**Figura 24**). Teniendo en cuenta los resultados en relación a los efectos de secuencia en los orígenes artificiales mostrados en el capítulo anterior, se trató de minimizar el número de sustituciones nucleotídicas para evitar en la medida de lo posible alteraciones en la actividad del origen. Por otra parte, con el fin de no interferir en las primeras posiciones, donde los efectos de secuencia son más acusados, y para facilitar una cierta distancia para un correcto posicionamiento del dominio de unión a DNA de GAL4, se realizaron los cambios a partir de la posición 12 (**Figura 24**).

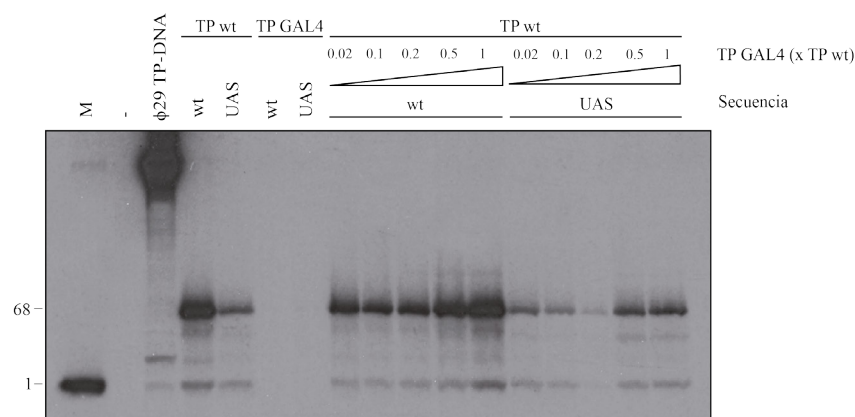


**Figura 24. Diseño y mecanismo de activación propuesto de la proteína TP GAL4.** Se diseñó la proteína TP GAL4 mediante la sustitución del dominio N-terminal de la TP del bacteriófago  $\phi 29$  por el dominio de unión a DNA del factor de transcripción GAL4. La construcción y purificación se describe en *Materiales y Métodos*. Para el estudio de la actividad de la proteína TP GAL4 se diseñó un origen artificial dsDNA de 68 pb al que se le introdujo la secuencia consenso de reconocimiento del dominio de unión a DNA de GAL4 (UAS, Upstream Activation Sequence). Las sustituciones en la secuencia se hicieron minimizando el número de cambios necesarios y teniendo en cuenta la distancia al extremo. Los nucleótidos cambiados se muestran en color rojo.

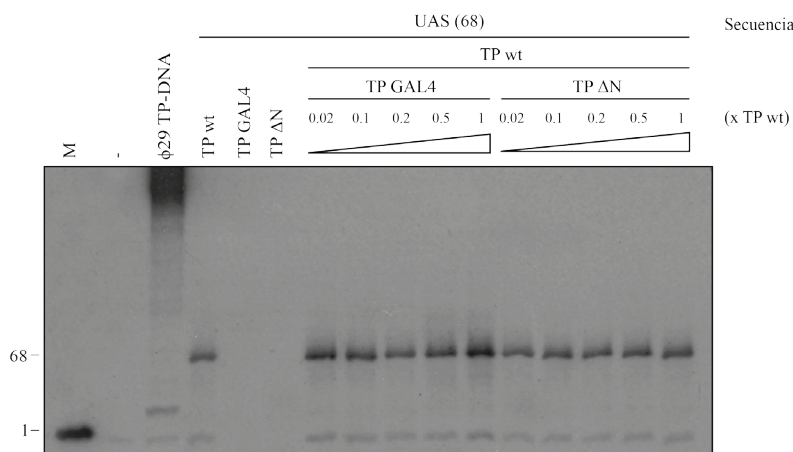
## 2. Utilización del origen artificial con la secuencia UAS con la proteína TP GAL4

El estudio sobre la utilización del origen con la secuencia UAS y la proteína TP GAL4 se llevó a cabo en un experimento de replicación estándar de DNA empleando [ $\alpha^{32}$ P]dATP como se describe en *Materiales y Métodos*. En éste se analizaron las proteínas TP wt y TP GAL4 a una concentración de 150 nM con los orígenes de 68 pb que tenían la secuencia *wild-type* (carriles wt) y la que contenía la UAS (carriles UAS). Los resultados observados en la **Figura 25A** muestran cómo la proteína TP wt es capaz de replicar ambos orígenes (carriles TP wt: wt y UAS) pero con una eficiencia unas 10 veces menor en el caso del origen que tiene la secuencia UAS, lo que resultó inesperado. Por otra parte, con la proteína TP GAL4 no se observa actividad replicativa con ninguno de los moldes. Aunque la proteína TP GAL4 parecía estar afectada en su actividad de replicación, en el caso en que existiera una interacción específica de GAL4 con la secuencia de reconocimiento UAS, la TP GAL4 podría reclutar por interacción entre TPs a la TP wt al origen, estableciendo así un efecto cooperativo en la unión al extremo. Para analizar esto, se rea-

A



B



**Figura 25. (A) Utilización del origen artificial con la secuencia UAS con la TP wt y TP GAL4.** El uso de los orígenes artificiales de 68 pb con la secuencia *wild-type* y UAS se analizó con las proteínas TP wt y TP GAL4 en un experimento de replicación como se describe en *Materiales y Métodos*, empleando [ $\alpha^{32}$ P]dATP como nucleótido marcado. Se indican las posiciones correspondientes con la iniciación TP-dAMP [+1] y longitud completa [+68] (TP-68 nucleótidos ssDNA). Carril M, marcador de iniciación (empleando el TP-DNA de  $\phi$ 29 con dATP como único nucleótido); carril  $\phi$ 29 TP-DNA, control de replicación (empleando TP-DNA de  $\phi$ 29 como molde y todos los nucleótidos). Las proteínas TP wt y TP GAL4 se ensayaron a una concentración de 150 nM con ambos moldes de DNA (carriles TP wt, wt y UAS, respectivamente; carriles TP GAL4, wt y UAS, respectivamente). A una concentración fija de 75 nM de TP wt, se ensayaron concentraciones crecientes de TP GAL4 (de 0.02 a 1 veces la concentración de TP wt) tanto con el molde wt como UAS (carriles 0.02-1, wt y UAS, respectivamente). **(B) Comparación en el uso del origen artificial con la secuencia UAS de la TP GAL4 y TP  $\Delta$ N.** La comparación en el uso del origen UAS entre la TP GAL4 y TP  $\Delta$ N se realizó mediante un experimento de replicación como se describe en *Materiales y Métodos*, empleando [ $\alpha^{32}$ P]dATP como nucleótido marcado. Se indican las posiciones correspondientes con la iniciación TP-dAMP [+1] y longitud completa [+68] (TP-68 nucleótidos ssDNA). Carril M, marcador de iniciación (empleando el TP-DNA de  $\phi$ 29 con dATP como único nucleótido); carril  $\phi$ 29 TP-DNA, control de replicación (empleando TP-DNA de  $\phi$ 29 como molde y todos los nucleótidos). Las proteínas TP wt, TP GAL4 y TP  $\Delta$ N se ensayaron a una concentración de 75 nM con el origen UAS (carriles TP wt, TP GAL4 y TP  $\Delta$ N). A una concentración fija de 75 nM de TP wt, se ensayaron concentraciones crecientes de TP GAL4 y TP  $\Delta$ N (de 0.02 a 1 veces la concentración de TP wt) (carriles TP GAL4, 0.02-1; carriles TP  $\Delta$ N, 0.02-1; respectivamente).

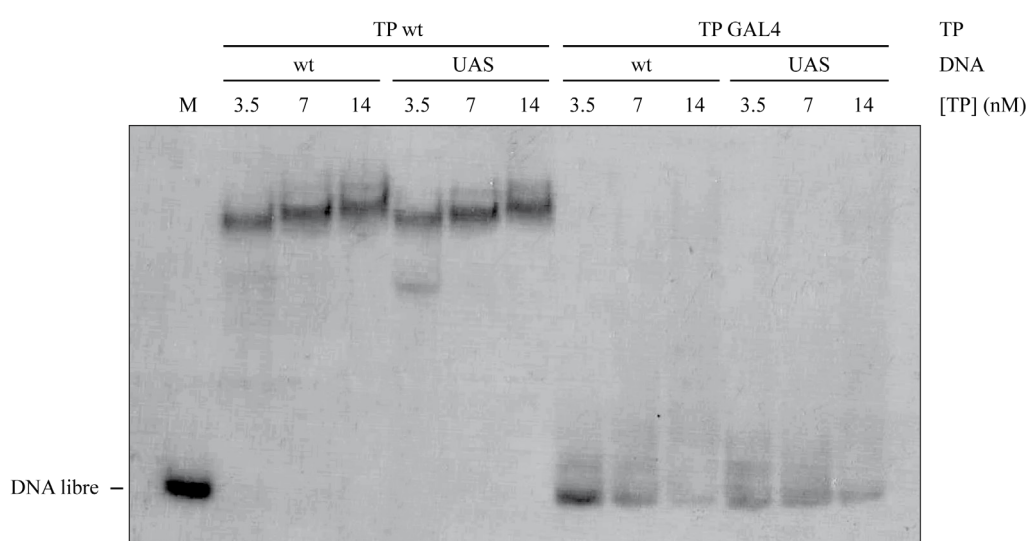
lizó un experimento de replicación con ambos moldes, manteniendo una concentración fija de proteína TP wt (75 nM) y aumentando progresivamente la de TP GAL4 (0.02-1 veces la concentración fija de TP wt). De esta forma, en el punto en el que hay una relación 1:1, hay la misma cantidad de TP que en el experimento estándar llevado a cabo con la TP wt (carril TP wt: wt y UAS). En ambos moldes (wt y UAS), se observa un comportamiento similar, donde la cantidad de producto disminuye conforme aumenta la cantidad de TP GAL4 para luego aumentar posteriormente hasta un nivel superior al inicial. Este comportamiento bifásico es difícil de explicar, pero parece que un aumento en la cantidad final de TP supone un aumento en la cantidad de producto sintetizado independientemente del molde empleado. Este efecto podría deberse a la ausencia del dominio N-terminal de la TP GAL4 o a la inclusión del dominio de unión a DNA del factor de transcripción GAL4. Para analizarlo, se llevó a cabo el mismo experimento utilizando el mutante de delección TP  $\Delta$ N (**Figura 25B**). En primer lugar, se observó la actividad de cada una de las proteínas (carril TP wt, TP GAL4 y TP  $\Delta$ N) a una concentración de 75 nM con el origen que tenía la secuencia UAS. Mientras que con la TP wt se obtenía una cantidad de producto moderada, con las proteínas TP GAL4 y TP  $\Delta$ N no se observó actividad. Por otro lado, sobre una concentración fija de 75 nM de TP wt se añadieron concentraciones crecientes de las proteínas TP GAL4 y TP  $\Delta$ N (0.02-1 veces la concentración fija de TP wt). En el caso de la TP GAL4 se observa un comportamiento similar al descrito más arriba (carriles TP GAL4, 0.02-1). Sin embargo, con la TP  $\Delta$ N no había variaciones en la cantidad de producto de replicación obtenido (carriles TP  $\Delta$ N, 0.02-1), ni en la concentración mayor ensayada, por lo que el efecto observado con la proteína TP GAL4 parecía ser consecuencia del dominio de unión a DNA de GAL4.

### 3. Capacidad de unión de la proteína TP GAL4 al DNA

Para analizar la capacidad de unión de la proteína TP GAL4 al DNA y su especificidad por la secuencia UAS introducida en el origen artificial de 68 pb de  $\phi$ 29 se realizó un experimento de retraso en gel en comparación con la TP wt. Para ello se utilizaron como sustrato los orígenes artificiales wt y UAS de 68 pb marcados radiactivamente en 5' (ver *Materiales y Métodos*) y se añadieron concentraciones crecientes de la TP wt o TP GAL4.

La **Figura 26** muestra que la TP wt forma un complejo estable con ambos orígenes con la aparición de bandas correspondientes a un retraso en la movilidad de la sonda de DNA marcada. Según aumenta la concentración de TP wt se observa una diferente movilidad de las bandas, lo que indicaría que, a medida que aumenta el número de moléculas de proteína, éstas se unen a varios sitios en el sustrato. Por el contrario, la proteína TP GAL4 se une de una forma muy poco eficiente a ambos sustratos, ya que no se observa

una banda concreta correspondiente al sustrato retrasado en ninguna de las concentraciones de proteína ensayadas. En su lugar, aparece un rastro difuso de producto a lo largo de todo el carril. Como se ha mencionado anteriormente, se sabe que la delección del dominio N-terminal elimina la capacidad de unión a dsDNA de la TP. Por ello, aunque se observa una gran disminución de la banda correspondiente al DNA libre a una concentración de proteína de 14 nM, el resultado indica que la inserción del dominio de unión a DNA de GAL4 no ha dotado a la fusión con TP de capacidad específica de unión a la UAS. Esto podría explicar la falta de actividad observada con esta proteína.



**Figura 26. Análisis de la unión a los orígenes de 68 pb de la proteína TP GAL4 mediante ensayos de retraso en gel.** Experimento de retraso en gel utilizando 1 nM de los orígenes artificiales marcados radiactivamente en 5' con [ $\gamma^{32}$ P]ATP con la secuencia del extremo derecho del genoma de  $\phi 29$  (carriles wt) o con la inserción de la secuencia consenso UAS de GAL4 (carriles UAS). Los sustratos se incubaron en ausencia (carril M, DNA libre) o presencia de concentraciones crecientes de la TP *wild-type* (carriles TP wt) o de la TP GAL4 (carriles TP GAL4). Las muestras se procesaron como se describe en *Materiales y Métodos*.



## Capítulo 4

**Identificación de sitios potenciales de inserción de secuencias peptídicas permitidos en la TP del bacteriófago  $\phi$ 29**





Durante la pasada década, diversas estrategias han sido desarrolladas para el cribado de librerías polipeptídicas en busca de propiedades específicas. Estas características distintivas han sido detectadas exitosamente en un proceso denominado *biopanning* basado en la asociación de la proteína a su ácido nucleico codificante (fenotipo-genotipo) y que incluye, entre otras, la técnica de DNA display (revisado en Sergeeva *et al.*, 2006). En este sentido, la unión del genoma de  $\phi 29$  a la TP, codificada por el mismo, hace del sistema del bacteriófago  $\phi 29$  una plataforma muy interesante para este tipo de tecnologías de cribado, pero también para otras aplicaciones como la transferencia génica o como complemento a la tecnología de aptámeros de DNA, entre otras. Esencialmente, la TP podría servir como el punto de anclaje para péptidos o proteínas que fueran codificadas por el propio DNA al que estuvieran unidas y se multiplicasen por el proceso de replicación tras una fase de selección. Sin embargo, hasta el momento no ha habido ningún estudio que permita encontrar sitios potenciales de inserción para la inclusión de secuencias peptídicas en una proteína como la TP sin comprometer su estructura-actividad.

## 1. Aproximación a la identificación de sitios potenciales de inserción permitidos en la secuencia de la TP

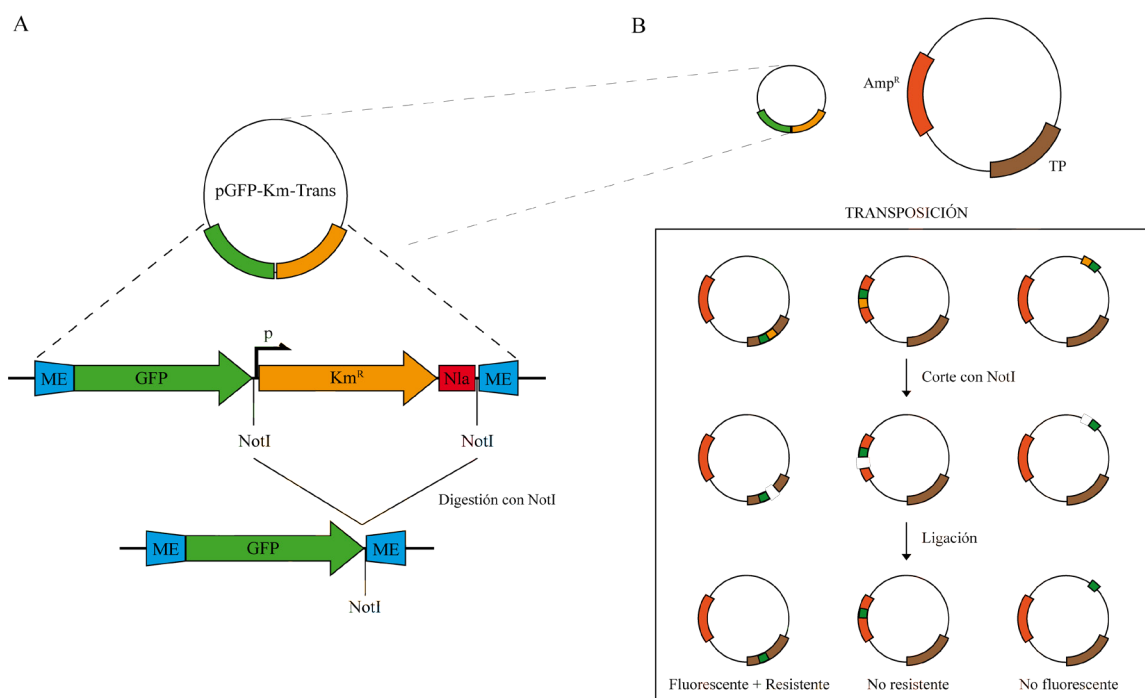
Recientemente se ha secuenciado el genoma del bacteriófago MG-B1 que infecta a la bacteria *B. weihenstephanensis* y codifica una putativa TP, además de una polimerasa y otras proteínas de replicación (Redondo *et al.*, 2013). El alineamiento de la secuencia de aminoácidos de esta TP con la de  $\phi 29$ , así como con un representante de cada grupo comprendido en el género de fagos  $\phi 29$ -like (PZA, grupo I; Nf, grupo II; y GA-1, grupo III), ha mostrado la presencia de dos inserciones principales en su secuencia de 12 y 22 aminoácidos, localizadas en las posiciones 142 y 158, respectivamente, separadas por 4 aminoácidos semiconservados en los tres grupos (**Figura 27**). La existencia de ambas inserciones sugiere que ciertas inserciones en la secuencia de la TP de  $\phi 29$  podrían ser permitidas. Como se ha mencionado en la *Introducción*, la TP presenta tres dominios: el dominio iniciador (C-terminal), el dominio intermedio y el dominio N-terminal. Salvo que el bucle iniciador que contiene la Ser232 es sensible a mutaciones, prácticamente nada se conoce en cuanto a los sitios de inserción de secuencias permitidos. Se ha observado que fusiones proteicas y peptídicas (tags) al dominio N-terminal son funcionales *in vivo* y en experimentos de actividad *in vitro* (Redrejo-Rodríguez *et al.*, 2013; Holguera *et al.*, 2015). La inserción de dominios proteicos en el interior de la secuencia de la TP supone ventajas potenciales como la posible simplificación de la propia TP o la eliminación, tras su uso al generar TP-DNA, de partes no útiles de la proteína para posteriores aplicaciones. Puesto que la búsqueda de estos sitios mediante un diseño racional puede ser un proceso muy complicado e incierto, se consideró un

100

buen abordaje la transposición al azar de un dominio proteico en la secuencia de la TP. Se ha descrito que para la efectividad de dicho proceso, el dominio a transponer debe cumplir con dos características: (i) tener los extremos N y C terminales próximos entre sí, para no causar una gran alteración en la proteína diana, y (ii) debe tener una estructura que le permita plegarse de forma autónoma (Krishna y Englander, 2005). La proteína verde fluorescente (GFP), de aproximadamente 30 KDa, cumple con las dos premisas anteriores, con los extremos amino y carboxilo terminales muy próximos entre sí y con un plegamiento, independiente de otras proteínas o cofactores, en forma de barril beta (revisado en Chalfie, 1995). De hecho, la GFP es un fluoróforo en sí misma, por lo que la emisión de fluorescencia es un indicador de un correcto plegamiento (Chalfie *et al.*, 1994). Desde el descubrimiento de su plegamiento independiente y fluorescencia, la GFP se ha empleado extensivamente en el diseño de proteínas de fusión con diversas aplicaciones (Wang y Hazelrigg, 1994; Marshall *et al.*, 1995). En la mayor parte de los casos, para evitar alteraciones en el plegamiento de la proteína diana, las fusiones se han realizado en los extremos N o C terminales. Sin embargo, para los casos en los que la actividad de estas proteínas con la fusión en los extremos se ve afectada o para abordajes como el propuesto aquí, se han diseñado transposones sintéticos que producen proteínas fusionadas a GFP en medio de su secuencia (Ross-Macdonald *et al.*, 1997, Merkulov y Boeke, 1998). Por otra parte, el transposón Tn5 ha demostrado ser una potente herramienta en la inserción de GFP en la secuencia de una proteína diana (Sheridan *et al.*, 2002). El transposón Tn5 consiste en cualquier secuencia flanqueada por unas ITR de 19 pb denominadas extremos mosaico (ME), los cuales, junto con la optimización de la transposasa, resultan en una frecuencia de inserción *in vitro* de un 1%. El mecanismo de acción supone la unión de la transposasa recombinante Tn5 a los ME, catalizando su inserción al azar en presencia de  $Mg^{2+}$  y produciendo un corte en el DNA diana que genera extremos protuberantes de 9 nucleótidos a ambos lados. Al ser reparados estos extremos, se produce la duplicación de esos 9 nucleótidos a ambos lados del transposón (revisado en Reznikoff *et al.*, 1999). Esta duplicación implica la incorporación de unos conectores relativamente cortos de 3 aminoácidos entre el transposón y la proteína en la que se aloja a cada lado (Biondi *et al.*, 1998; Hughes *et al.*, 2001). Por todo lo anterior, se optó por la transposición al azar de GFP en la TP de  $\phi 29$  basada en el transposón Tn5.

## 2. Transposición y selección de proteínas

Para la transposición se empleó como DNA donador un plásmido cedido por Belén Calles y Víctor de Lorenzo que contenía, entre dos ME, el gen de la GFP sin promotor ni RBS junto al gen de resistencia a kanamicina ( $Km^R$ ) regulado por su propio promotor (Billerbeck *et al.*, 2013). El gen de resistencia estaba flanqueado por dos dianas para el enzima de restricción NotI, por lo que el cas-



**Figura 28. Construcción de la librería de transposición.** (A) Esquema del plásmido donador que contenía el transposón diseñado para la transposición aleatoria en la secuencia de la TP de  $\phi 29$ . El transposón contenía, entre los dos extremos MOSAICO (ME), el gen de la GFP sin promotor regulador de su transcripción ni RBS junto al gen de resistencia a kanamicina ( $Km^R$ ), regulado por su promotor. El gen de resistencia estaba flanqueado por dos dianas para el enzima de restricción NotI, por lo que podía ser liberado mediante su corte y posterior ligación. (B) Esquema del proceso de creación de la librería de transposición (véase texto y *Materiales y Métodos* para una explicación detallada).

sete de resistencia podía ser liberado mediante corte y posterior ligación, dejando únicamente la secuencia codificante de la GFP en fase con los extremos del transposón (**Figura 28A**). El DNA diana contenía el gen de la TP de  $\phi 29$  con un doble Strep-tag fusionado al extremo N-terminal. La transposición se realizó como se describe en *Materiales y Métodos*. Los plásmidos que contenían el transposón se seleccionaron con kanamicina y ampicilina. Mediante la digestión con el enzima NotI y posterior religación, se obtuvo la librería NotI: el conjunto de plásmidos que tenían insertada una GFP en su secuencia (exentos del gen de resistencia a kanamicina como se ha explicado anteriormente). La inserción del transposón Tn5 es principalmente aleatoria, por lo que el transposón puede ser insertado con igual probabilidad en cualquier posición del plásmido. Así, éste podía insertarse en DNA no codificante (sin fluorescencia), en algún elemento fundamental para la replicación del plásmido o desarrollo de la resistencia al antibiótico (no se generan clones, si la función se ve afectada por la inserción) o en la región codificante de la TP (fluorescente y resistente, si el transposón se inserta en la orientación y un marco de lectura correctos).

De tal forma, la probabilidad de que el transposón se inserte en la secuencia de la TP es la relación entre el tamaño de la secuencia de la propia TP y la total del plásmido (27.5%). Sin embargo, la probabilidad de que el transposón se inserte en elementos fundamentales para la supervivencia de las bacterias, como el origen de replicación o la resistencia a ampicilina, hace que la probabilidad ascienda hasta un 47.4% para inserciones en la TP en las bacterias que sobreviven. Por otro lado, únicamente la mitad de ellas tendrán el transposón correctamente orientado (23.7%) y únicamente un tercio de estas se insertará en el marco de lectura correcto (7.9%) (**Figura 28B**). Por esa razón, se generó una librería de 5000 clones para obtener alrededor de 200 inserciones en la TP.

Placa I

	- GFP	1	2	3	4	5	6	7	8	9	10	+ GFP
A	374	64	61	81	64	64	66	67	65	68	60	325
B	342	60	72	63	71	62	85	66	62	62	68	883
C	346	63	62	63	66	67	66	69	65	76	77	1177
D	365	64	65	81	71	71	70	67	65	78	86	790
E	333	72	68	66	66	77	72	67	66	61	64	731
F	353	61	60	82	82	73	71	66	79	65	60	351
G	332	65	65	63	66	71	75	64	59	62	74	357
H	359	59	64	60	67	67	63	67	63	64	60	287
Media	351	613 (100%)										

Placa II

	- GFP	1	2	3	4	5	6	7	8	9	10	+ GFP
A	362	53	53	58	57	54	55	58	53	51	65	1290
B	380	54	53	58	65	52	53	53	65	52	58	1193
C	353	65	54	52	51	60	61	57	52	56	50	1119
D	345	51	53	62	55	61	56	73	54	51	55	667
E	343	52	53	59	54	60	60	48	55	50	50	636
F	357	50	55	53	55	55	57	50	117	53	50	415
G	360	51	55	57	54	61	56	51	40	50	50	420
H	366	49	52	54	53	53	56	55	52	53	56	266
Media	358	751 (100%)										

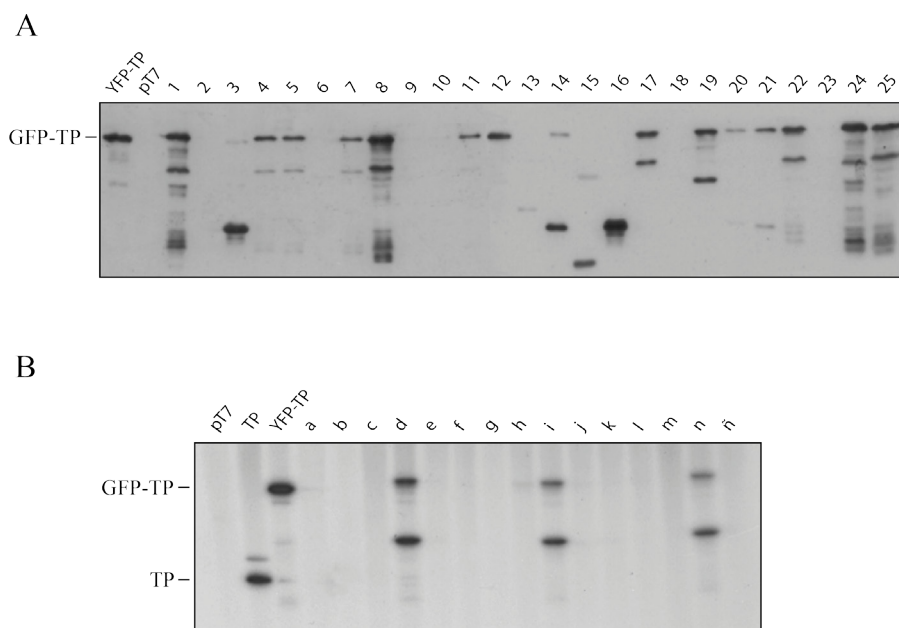
Placa III

	- GFP	1	2	3	4	5	6	7	8	9	10	+ GFP
A	363	61	63	70	71	60	63	67	57	58	52	1004
B	390	59	59	56	55	65	66	60	57	54	48	941
C	387	60	82	66	60	67	69	58	56	54	60	1113
D	383	61	76	66	69	70	70	72	57	54	66	762
E	375	57	64	69	60	77	65	66	56	56	59	653
F	370	61	62	58	62	62	65	56	53	57	47	374
G	344	58	56	50	56	54	59	57	54	61	62	401
H	378	61	57	56	53	61	57	61	53	49	49	357
Media	374	701 (100%)										

Placa IV

	- GFP	1	2	3	4	5	6	7	8	9	10	+ GFP
A	359	57	60	54	53	51	73	56	61	56	56	925
B	351	52	56	52	55	54	57	55	55	61	54	1012
C	424	54	57	55	57	57	64	57	63	58	58	1049
D	378	73	55	59	57	61	67	54	63	60	59	755
E	408	60	56	58	59	58	58	57	50	53	56	681
F	362	62	54	54	67	52	54	55	57	78	60	374
G	400	60	55	64	61	51	61	60	55	68	51	475
H	337	52	53	74	54	65	53	67	58	57	52	258
Media	377	691 (100%)										

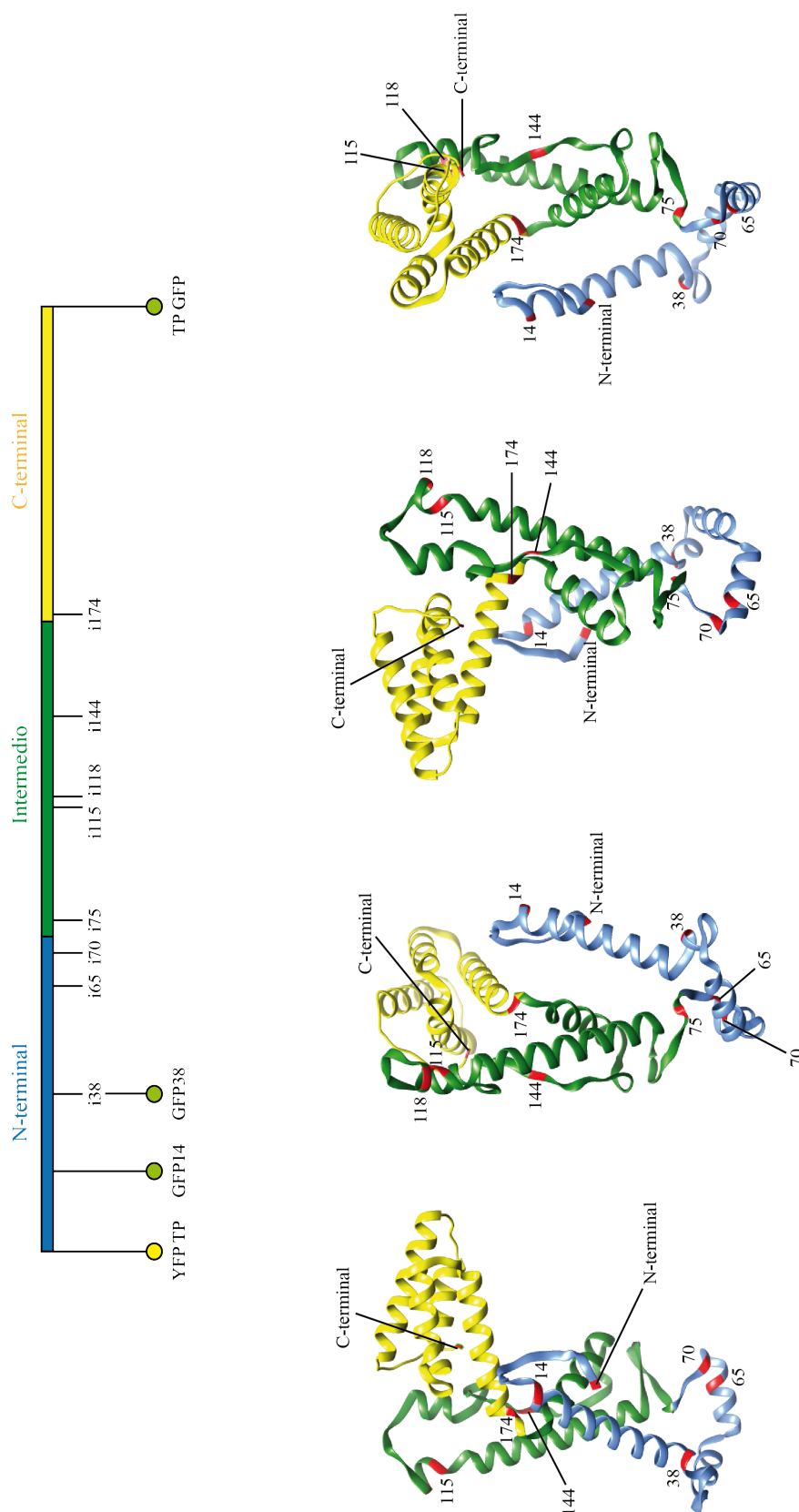
**Figura 29. Análisis de la fluorescencia emitida por los cultivos bacterianos que expresan la librería de transposición.** Se analizaron en placas de 96 pocillos y empleando un lector de placas OPTIMA (BMG LABTECH) un total de 320 clones. La columna indicada como -GFP corresponde a las medidas de bacterias que no expresaban GFP (control -). La columna indicada como +GFP corresponde a las medidas de bacterias que expresaban GFP (control +). Para analizar el valor estándar de fluorescencia se realizó la media de la columna +GFP y se le asignó un porcentaje del 100%. Los valores de fluorescencia de las proteínas experimentales obtenidos se expresan en porcentajes con respecto a dicho 100%, seleccionándose un total de 53 (marcadas en verde).



**Figura 30. (A) Análisis de la presencia de GFP en los extractos bacterianos que expresan la librería de transposición.** La detección de la presencia de GFP inserta en la secuencia de las distintas TPs expresadas se realizó mediante *western blot* con anticuerpos específicos anti-TP. Los carriles 1-25 muestran una representación de todas las construcciones analizadas. **(B) Análisis de la formación del complejo TP-dAMP en los extractos bacterianos que expresan la librería de transposición.** La reacción de iniciación en extractos bacterianos empleando TP-DNA como molde se realizó y procesó como se describe en *Materiales y Métodos* empleando [ $\alpha^{32}$ P]dATP como nucleótido marcado. En el experimento se utilizaron 35 nM de DNA polimerasa, 1 nM TP-DNA y se incubó durante 5 minutos a 30°C. Carril pT7, control negativo de extracto bacteriano que contenía el plásmido pT7 sin el gen de la TP. Carril TP, control positivo de extracto bacteriano que contenía el plásmido pT7 con el gen de la TP. Carril YFP-TP, control positivo de extracto bacteriano que contenía el plásmido pT7 con el gen de la TP con la proteína YFP fusionada en su extremo N-terminal. Carriles a-n, distintas fusiones TP-GFP analizadas. Las bandas inferiores en los carriles d, i y n corresponden a productos de proteólisis.

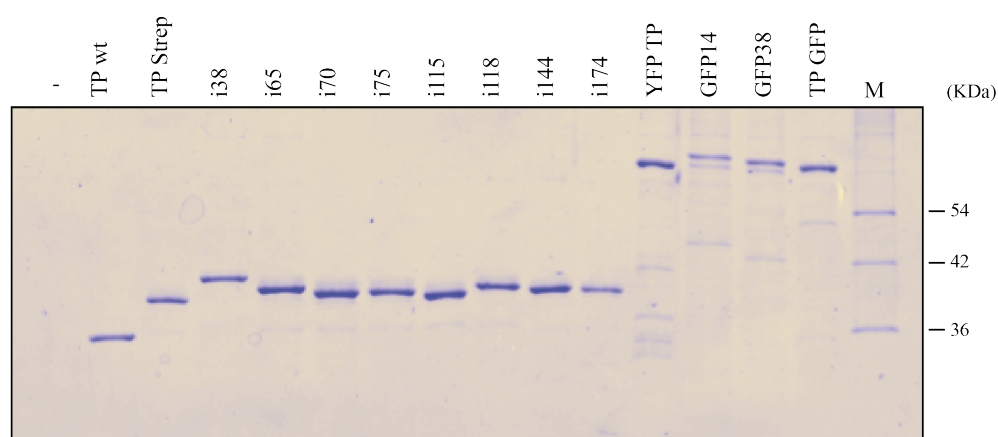
Las proteínas se seleccionaron en base a dos criterios: (i) fluorescencia emitida en cultivos bacterianos que expresaban las proteínas en placas de 96 pocillos (**Figura 29**); y (ii) posterior detección de GFP fusionada a la TP por *western blot* (**Figura 30A**) (véase *Materiales y Métodos*). Los clones positivos se cultivaron y se aisló su DNA plasmídico para su análisis por restricción (para determinar la posición de la GFP) y posterior secuenciación. Así se seleccionaron los clones con inserciones repartidas por la secuencia de la TP. Los clones seleccionados fueron sometidos a experimentos de iniciación de la replicación en extractos bacterianos empleando como molde el TP-DNA de  $\phi 29$  (**Figura 30B**). Este experimento es muy específico y permite detectar la función de TP sin necesidad de purificar la proteína. La mayor parte de las TPs que contenían la GFP inserta en su secuencia no resultaron activas en los experimentos de iniciación en extractos. Puesto que la presencia de una inserción del tamaño de la GFP podía inactivar la TP, decidimos analizar la permisividad en esas posiciones a inserciones cortas. Para ello se redujo su tamaño mediante una amplifi-





**Figura 31. Localización de las GFP o inserciones de 15 aminoácidos en el mapa y estructura de la TP.** Las inserciones de 15 aminoácidos se nombran con la letra i (inserción) seguido del número correspondiente a su posición dentro de la cadena polipeptídica. Las inserciones de GFP se nombran como sigue: YFP TP, proteína YFP fusionada al extremo N-terminal de la TP; GFP14 y GFP38, proteína GFP inserta en las posiciones 14 y 38 de la cadena polipeptídica, respectivamente; y TP GFP, proteína GFP fusionada al extremo C-terminal de la TP. El modelado tridimensional se realizó utilizando el programa UCSF Chimera (Pettersen *et al.*, 2004) a partir de los datos cristalográficos contenidos en el número de acceso 2EX3 de la Protein Data Bank (PDB) (Kamtekar *et al.*, 2006) y una reconstrucción del dominio N-terminal mediante el programa I-Tasser (Roy *et al.*, 2010). En el modelado se muestra en color rojo la localización de las posiciones que contienen la fusión a GFP/YFP o las inserciones. El código de colores para los diferentes dominios es el mismo que el empleado en el mapa superior.

cación de los respectivos plásmidos con los oligonucleótidos GFPup y GFPdown (véase *Materiales y Métodos*). Estos oligonucleótidos hibridan en el extremo N-terminal y C-terminal de la GFP hacia el exterior de su secuencia, de forma que su uso en una reacción de amplificación con los plásmidos que contenían el transposón eliminaban la secuencia de la GFP y dejaban en su lugar una inserción más pequeña correspondiente a 15 aminoácidos y secuencia peptídica: GVDLSLTAADVYKRQ. Al análisis posterior de estas proteínas se añadieron las proteínas YFP TP y TP GFP que se construyeron mediante la fusión de la YFP en el extremo N-terminal y la GFP en el C-terminal, respectivamente (véase *Materiales y Métodos*). La **Figura 31** muestra el mapa de la localización de las fusiones a GFP o de los insertos reducidos de 15 aminoácidos en las proteínas que fueron purificadas para su ulterior análisis bioquímico. El resultado de las purificaciones de las proteínas se muestra en la **Figura 32**.



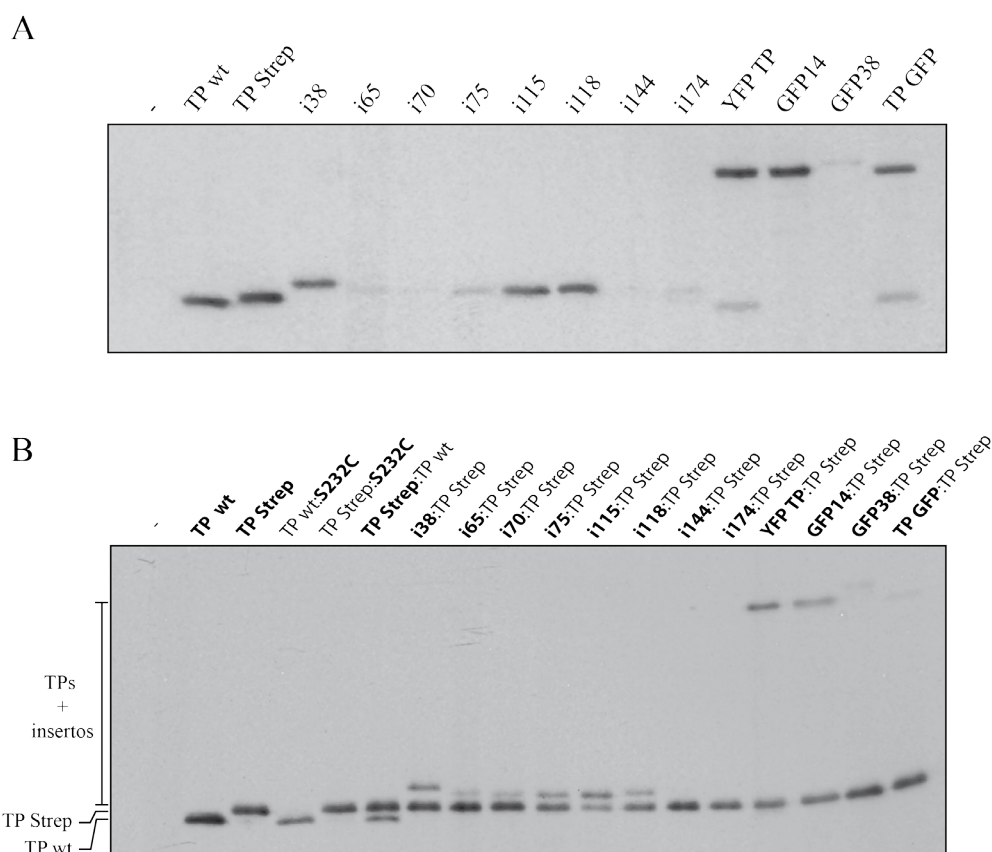
**Figura 32. Purificación de las TPs seleccionadas.** Las proteínas se purificaron como se describe en *Materiales y Métodos* y se analizaron por SDS-PAGE y posterior tinción con azul de Coomassie. La nomenclatura es la que se describe en la **Figura 31**. Como marcador de peso molecular se empleó el mutante *sus14* del bacteriófago  $\phi 29$  y se muestran las bandas correspondientes a las proteínas p10 (36 KDa), p11 (42 KDa) y p8 (54 KDa) (carril M).

### 3. Iniciación de las TP con insertos y competición por la DNA polimerasa

Para observar el efecto que tenían las diferentes inserciones en los primeros eventos de la replicación del TP-DNA, se llevaron a cabo experimentos de iniciación con las TPs con insertos, empleando las proteínas purificadas y  $[\alpha^{32}\text{P}]\text{dATP}$ , y se compararon con la actividad de la TP wt. Como se observa en la **Figura 33A**, la TP que tenía un doble Strep-tag en el extremo N-terminal (TP Strep) mostraba una eficiencia de iniciación similar a la TP wt. La TP Strep es la actividad *wild-type* a la que referiremos la actividad



de las TPs con insertos, pues todas llevan el mismo tag. Las proteínas YFP TP y GFP14 mostraron una actividad similar a la TP Strep. Las proteínas i38, i115, i118 y TP GFP mostraron una actividad ligeramente inferior a la TP wt. Por otra parte, las demás proteínas mostraban una actividad muy baja o no detectable.



**Figura 33. (A) Formación del complejo TP-dAMP con la TP wt y las TPs con inserciones empleando TP-DNA como molde.** La reacción de iniciación se realizó y procesó como se describe en *Materiales y Métodos* empleando [ $\alpha^{32}$ P]dATP como nucleótido marcado. En el experimento se utilizaron 150 nM de TP wt o variantes, 30 nM de DNA polimerasa, 0.63 nM TP-DNA y se incubó durante 5 minutos a 30°C. Carril -, control negativo en ausencia de TP. La nomenclatura de las diferentes proteínas es la empleada anteriormente. La figura es representativa de, al menos, tres experimentos independientes. **(B) Experimento de competición entre la TP wt y las TPs con inserciones.** La reacción de competición se realizó y procesó como se describe en *Materiales y Métodos* empleando [ $\alpha^{32}$ P]dATP como nucleótido marcado. En el experimento se utilizaron 75 nM de cada una de las dos TP de la competición ( $[TP_{total}] = 150$  nM), 30 nM de DNA polimerasa, 0.63 nM TP-DNA y se incubó durante 5 minutos a 30°C. La DNA polimerasa se añadió después de poner en contacto ambas TPs. Como control de interferencia se utilizó una TP mutada en el residuo iniciador (S232C) que es prácticamente inactiva aunque conserva una interacción wild-type con la DNA polimerasa (Garmendia et al., 1990). La figura es representativa de, al menos, tres experimentos independientes.

Cabe destacar que la proteína i38, que contiene la inserción de 15 aminoácidos en la posición 38, realiza una iniciación muy eficiente en comparación con la que tiene inserta una GFP en la misma posición, lo que sugiere que la secuencia insertada probablemente cause una interferencia con el plegamiento y, por tanto, la función de la TP.

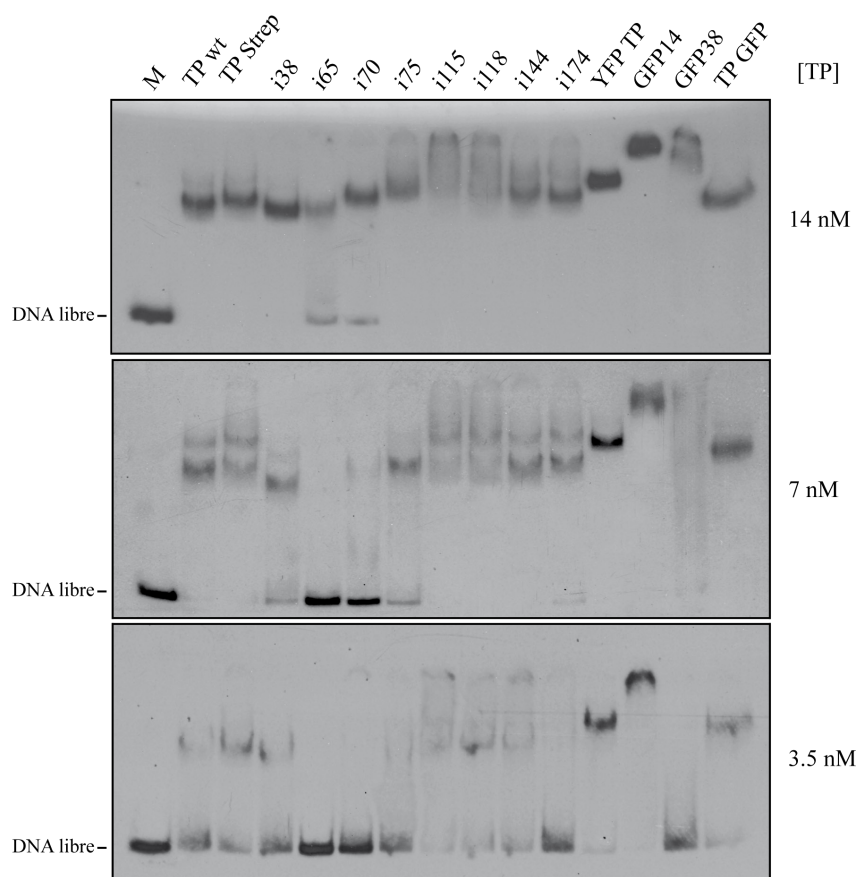
Para obtener una mejor comparación con la actividad de iniciación de la TP wt, se realizaron experimentos de interferencia, donde la interacción de la TP *wild-type* era competida por las TPs con insertos en un experimento de iniciación utilizando como molde TP-DNA (véase *Materiales y Métodos*). Este experimento evalúa una combinación de varias interacciones: (i) TP/DNA polimerasa; (ii) TP/DNA; y (iii) TP-TP parental. En primer lugar, la TP Strep se analizó en competición con la TP wt con el fin de observar si el Strep tag estaba afectando a la unión con la DNA polimerasa/TP-DNA. Puesto que las TPs que contenían insertos tenían un doble Strep-tag fusionado al extremo N-terminal, su competitividad se evaluó con respecto a la TP con Strep tag (TP Strep). Como control de interferencia se utilizó una TP mutada en el residuo iniciador (S232C) que es prácticamente inactiva en iniciación aunque conserva una interacción *wild-type* con la DNA polimerasa (Garmendia *et al.*, 1990). Se pusieron en contacto la misma cantidad de ambas TPs y posteriormente se añadió la DNA polimerasa para mantener la misma probabilidad en su contacto con ambas TPs. En la **Figura 33B** se observa que la TP Strep inicia aparentemente de forma más eficiente que la TP wt, puesto que en su competencia con el mutante S232C, la actividad de iniciación de la TP wt desciende un 70%, mientras que con la TP Strep desciende un 20%. De todas las proteínas ensayadas, la única que interacciona mejor que la TP Strep es la i115, con una actividad de iniciación un 5% superior a la TP Strep. Por otra parte, las proteínas i38, i75, i118, YFP TP y GFP14 oscilan en un rango de 30-70% de la actividad de iniciación obtenida con la TP Strep. Con las proteínas i65, i70, GFP38 y TP GFP se observan actividades de iniciación inferiores al 14% de la TP Strep y, como se había visto anteriormente, los mutantes i144 e i174 no inician. Esto concuerda con lo observado en la **Figura 33A**.

#### 4. Capacidad de unión a DNA de las proteínas con insertos

El análisis de la capacidad de unión a dsDNA de las proteínas con insertos se llevó a cabo mediante experimentos de retraso en gel (véase *Materiales y Métodos*). Como sustrato se empleó un fragmento de dsDNA de 216 pb correspondientes al gen *yshC* del genoma de *B. subtilis* marcado radiactivamente en el extremo 5'. El experimento se realizó a 3 concentraciones diferentes de proteína, 3.5, 7 y 14 nM. La **Figura 34** muestra que todas las proteínas producen un retraso en la banda correspondiente al sustrato de DNA marcado y son capaces, por tanto, de unirse a él. Sin embargo, en comparación con el resto, las proteínas i65 e i70 están moderadamente afectadas en su unión, ya que a la mayor concentración de proteína

ensayada (14 nM) aún se observa DNA libre, y las proteínas i38, i75 y GFP38 están levemente afectadas, ya que a la concentración intermedia de proteína ensayada (7 nM) aún se observa DNA libre. Estas proteínas afectadas en la unión a DNA corresponden a proteínas que tienen inserciones fundamentalmente en el dominio N-terminal (la inserción de la proteína i75, si bien estrictamente se encuentra en el dominio intermedio, es muy próxima al límite con el dominio N-terminal, situado en la posición 73). Gracias a estudios anteriores, se sabe que este dominio es el responsable de la unión a DNA inespecífica de secuencia que presenta la TP wild-type (Garmendia *et al.*, 1988) y se ha determinado la importancia de los residuos 14-38 en dicha unión (Redrejo *et al.*, 2012; Holguera *et al.*, 2014). Por ello, las inserciones en este dominio podrían haber afectado su capacidad de unión.

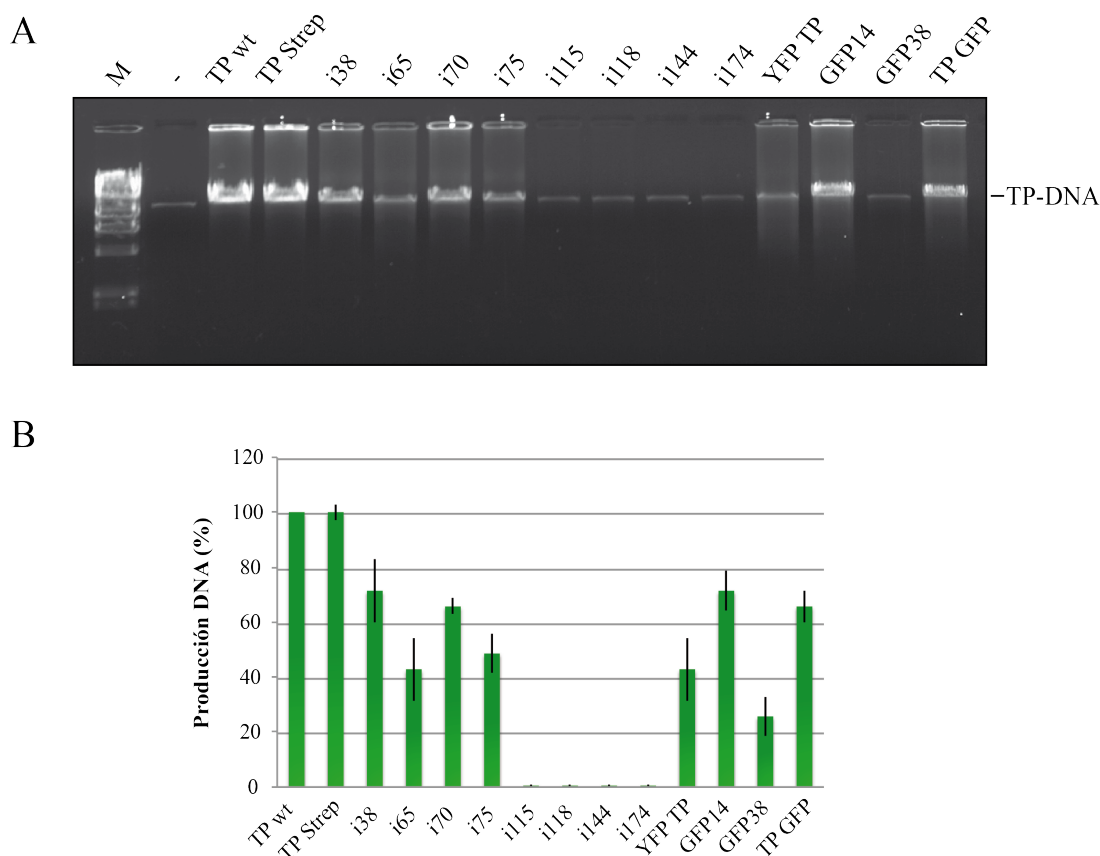
En general, a excepción de la proteína i38, parece que las inserciones en el dominio N-terminal que reflejaban problemas en la unión al molde de DNA están afectadas en su actividad de iniciación. Sin embargo, otras proteínas donde no se había visto un defecto en su capacidad de unión a DNA tampoco muestran actividad de iniciación como las proteínas i144 o i174.



**Figura 34. Análisis de la unión a dsDNA de las proteínas con insertos o GFP mediante ensayos de retraso en gel.** Experimento de retraso en gel empleando 1 nM de un fragmento de dsDNA de 216 pb (correspondiente al gen *yshC* del genoma de *B. subtilis*) marcado en el extremo 5' con  $[\gamma^{32}\text{P}]\text{ATP}$ . El sustrato marcado fue incubado en ausencia (carril M) o en presencia de diferentes concentraciones de las proteínas ensayadas (3.5, 7 y 14 nM), indicadas en cada caso. Se indica la banda correspondiente al DNA libre. El experimento se realizó como se detalla en *Materiales y Métodos*.

## 5. Amplificación del TP-DNA de $\phi 29$

Puesto que la amplificación de DNA cebada por proteína terminal supone un paso fundamental para posteriores aplicaciones, estas proteínas se ensayaron en experimentos de amplificación empleando el TP-DNA de  $\phi 29$  como molde y se analizaron mediante su detección en geles de agarosa por tinción con bromuro de etidio (**Figura 35**). En este caso, la cuantificación de las bandas se realizó mediante su análisis con el programa ImageJ, mediante el que se obtuvieron los datos relativos a la intensidad de la banda correspondiente al tamaño del TP-DNA. Los valores se estandarizaron tomando la cantidad de producto obtenida con la TP wt como el 100%.



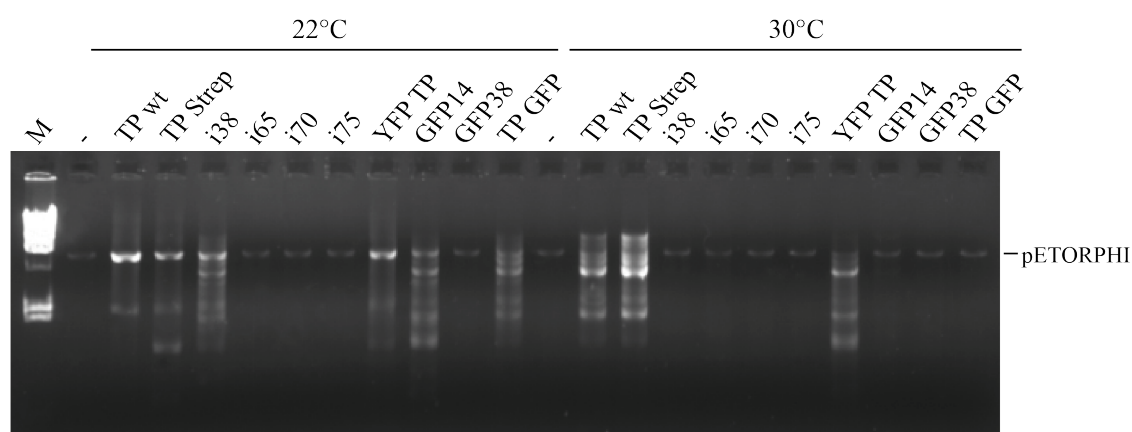
**Figura 35. Amplificación de TP-DNA con las TP con insertos.** La reacción de amplificación se realizó y procesó como se describe en *Materiales y Métodos*. En el experimento se utilizaron 6 nM de DNA polimerasa, 24 nM de las diferentes TPs, 35  $\mu$ M de p6, 30  $\mu$ M de p5 y se incubó durante 2 horas a 30°C. Carril M, marcador de peso molecular del DNA del fago  $\lambda$  digerido con HindIII; carril -, control negativo de la amplificación en ausencia de TP (DNA input). La nomenclatura de las diferentes proteínas es la empleada anteriormente. Se indica la posición correspondiente al TP-DNA. **(A)** El panel superior muestra el análisis de tamaño del DNA producido mediante la tinción con bromuro de etidio. **(B)** El gráfico de barras del panel inferior muestra la producción de TP-DNA para cada una de las distintas TPs tomando la producción de la TP wt como el 100%. Las barras se presentan como los valores medios y su desviación estándar, obtenidos a partir de, al menos, tres experimentos independientes

En este caso, con las proteínas TP wt y TP Strep se observó la misma eficiencia de amplificación, por lo que el Strep tag no parecía afectar a la actividad de la TP en amplificación, como se había visto previamente. Con las proteínas i38, i70, GFP14 y TP GFP se obtuvieron valores de 65-71% la actividad de la TP wt, lo cual concuerda con la actividad obtenida en los experimentos de iniciación para las proteínas i38, GFP14 y TP GFP. Con las proteínas i65, i75 e YFP TP se obtuvieron valores de 42-48% la actividad de la TP wt, lo cual podría parecer sorprendente porque con las proteínas i65 e i75, al igual que con la i70, se obtenían valores muy bajos en la actividad de iniciación. Sin embargo, la estimulación de la polimerización a consecuencia del incremento en 1000 veces la concentración de dNTPs empleada en los experimentos de amplificación con respecto a los de iniciación, podría ser la explicación para la recuperación de la actividad replicativa observada para dichas TPs. La proteína GFP38 resultó prácticamente inactiva. Por otra parte, con las proteínas i115, i118, i144 e i174 no se observó actividad de amplificación. Cabe destacar que, a excepción de i75, las proteínas con las que no se observó actividad de amplificación eran las que contenían inserciones localizadas en el dominio intermedio o en el C-terminal. Por otra parte, el descenso en la actividad observado con la GFP38 en relación a la i38, volvía a poner de manifiesto la importancia del tamaño de la secuencia insertada.

## 6. Amplificación de DNA heterólogo: pETORPHI

Las proteínas que habían resultado más activas en la amplificación con el TP-DNA de  $\phi 29$  (además de la GFP38 para su comparación con i38) se analizaron en experimentos de amplificación con el plásmido pETORPHI linealizado con DraI (véase *Capítulo 1*). La amplificación del pETORPHI supone el paso inmediatamente anterior al desarrollo de futuras aplicaciones, puesto que las proteínas que sean capaces de amplificarlo serán las que puedan ser empleadas en la fusión de nuevas funciones a DNAs heterólogos. El experimento se realizó como se describe en *Materiales y Métodos* a dos temperaturas diferentes: 22°C y 30°C (**Figura 36**). Como se había comprobado anteriormente (véase *Capítulo 1*), la temperatura óptima para la amplificación del pETORPHI resultó ser de 22°C, puesto que a 30°C se observaban productos de menor peso molecular. A 30°C sólo eran activas la TP wt, la TP Strep, ésta más eficiente incluso que la propia TP wt, y la YFP TP. Sin embargo, con ninguna de ellas se obtenía una única banda discreta correspondiente al tamaño del plásmido pETORPHI linealizado, sino un patrón de bandas heterogéneo cuyo producto mayoritario era de un tamaño inferior al unidad. A 22°C, la TP Strep resultó menos eficiente que la TP wt y el producto mayoritario obtenido con las proteínas i38, que era completamente inactiva a 30°C, e YFP TP era de un tamaño similar al obtenido con la TP Strep. En el caso de las proteínas GFP14 y TP GFP a 22°C se observó un patrón de bandas entre las que se encontraba la correspondiente a un producto

del tamaño del pETORPHI linealizado, pero no era la mayoritaria. Por otra parte, las proteínas i65, i70, i75 y GFP38 resultaron inactivas a ambas temperaturas, aun cuando todas ellas permitían la amplificación del TP-DNA de  $\phi 29$ . Los defectos de unión al DNA en un molde en el que no hay TP parental probablemente estén impidiendo su actividad.

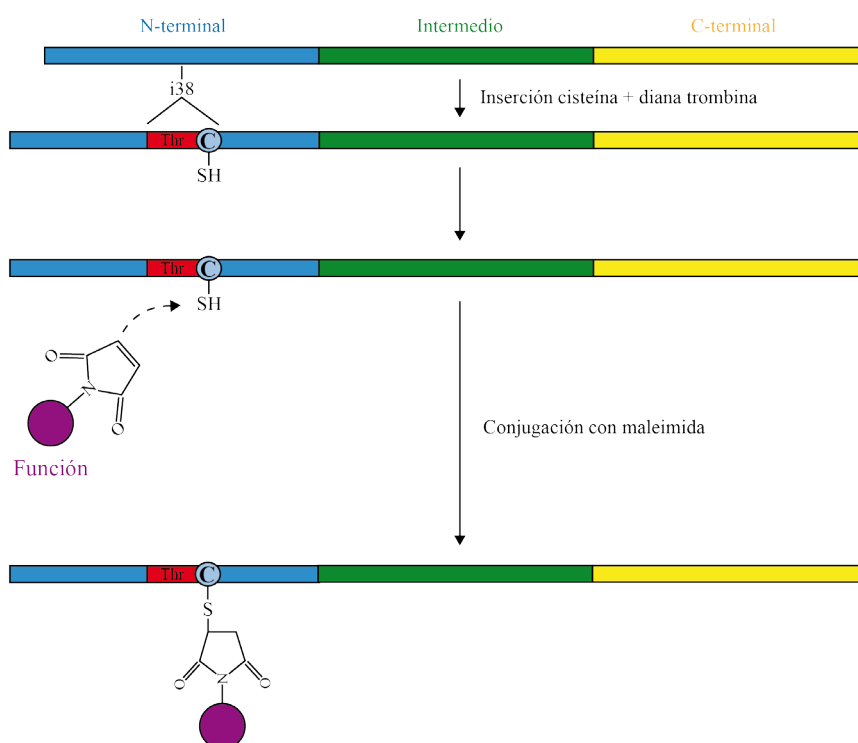


**Figura 36. Amplificación del pETORPHI con las TP con insertos.** La reacción de amplificación se realizó y procesó como se describe en *Materiales y Métodos*. En el experimento se utilizaron 15 nM de DNA polimerasa, 60 nM de las diferentes TPs, 5  $\mu$ M de p6, 30  $\mu$ M de p5 y se incubó durante 2 horas a las temperaturas indicadas. Carril M, marcador de peso molecular del DNA del fago  $\lambda$  digerido con HindIII; carril -, control negativo de la amplificación en ausencia de TP (DNA *input*). La nomenclatura de las diferentes proteínas es la empleada anteriormente. Se indica la posición correspondiente al TP-DNA. La figura muestra la fotografía del gel de agarosa teñido con bromuro de etidio y es representativa de, al menos, tres experimentos independientes.

## 7. La TP i38 y su aplicación biotecnológica

De todas las TPs analizadas en este estudio, la única que contenía una inserción peptídica en el interior de su secuencia y que era activa en amplificación con DNA heterólogo de forma comparable a la TP Strep era la i38. Como se ha comentado en la *Introducción*, se ha descrito que el fragmento mínimo de la TP para su unión a DNA consiste en los 70 primeros aminoácidos y que, con la proteína completa, los primeros 15 aminoácidos son prescindibles para dicha función. Además, el mutante K27A y el doble mutante K25A/K27A del dominio N-terminal de la TP son deficientes en la capacidad de unión a DNA (Holguera *et al.*, 2014). Por otra parte, se ha descubierto recientemente una señal de localización nuclear eucariótica (NLS) localizada en los primeros 37 aminoácidos del dominio N-terminal de la TP de  $\phi 29$  y

otros bacteriófagos (Redrejo-Rodríguez *et al.*, 2012; Redrejo-Rodríguez y Salas, 2014), que convierten al sistema de  $\phi 29$  en una herramienta potencial de translocación nuclear de genes en células de mamífero. Sin embargo, la capacidad de unión inespecífica de secuencia puede suponer un problema en la aplicación del sistema en los procesos de transferencia génica, así como la NLS, que restringe al núcleo el destino del DNA que va covalentemente unido a la TP. Por esa razón, la posibilidad de inserción de secuencias en la posición 38 supone una ventaja clave, ya que se podría insertar un grupo reactivo para la fusión de una nueva función y, simultáneamente, una diana para una proteasa que permitiese deletar el fragmento 1-38, con la consiguiente eliminación de la NLS y la función de unión inespecífica a DNA.



**Figura 37. Diseño de la proteína TPi38cys y mecanismo de conjugación con el grupo maleimida.** La proteína TPi38cys se construyó mediante la inserción de 17 aminoácidos (ASGGLVPRGSGGCGGGT) en la posición 38 de la TP de  $\phi 29$ . Esta inserción contenía una diana para la proteasa trombina y una cisteína cuyo grupo tiol (-SH) es capaz de reaccionar de forma espontánea con un grupo maleimida de la forma en que se muestra.



### 7.1. Diseño de la aplicación

Para probar el abordaje descrito en el epígrafe inmediatamente anterior, se clonó en la posición 38 de la TP una secuencia consistente en: dos glicinas, una diana para la proteasa trombina seguida de dos glicinas a modo de espaciador, una cisteína como grupo reactivo y otras dos glicinas a modo de espaciador (GGLVPRGSGGCGG). Junto con los dos aminoácidos a ambos lados de la inserción correspondientes a la traducción de las dianas de los enzimas de restricción empleados para su clonaje, en total se insertaron 17 aminoácidos en la posición 38. La cisteína aportaría el grupo reactivo –SH para la adición de nuevas funciones y la diana para la trombina permitiría la eliminación de los primeros 38 aminoácidos de la TP en caso de que fuera necesario. A esta proteína se la denominó TPi38cys. Para el análisis de la potencia- lidad de este abordaje, se optó por fusionar a la TP dos péptidos de penetración celular (péptidos TAT: AGYGRKKRRQRRR y Antennapedia (ANTP): SGRQIKIWFQNRRMKWKK) y se empleó como grupo reactivo una maleimida, que reacciona de forma espontánea con el grupo tiol de la cisteína. El diseño de la proteína TPi38cys así como el mecanismo de conjugación por medio de un grupo maleimida se ilustra en la **Figura 37**.

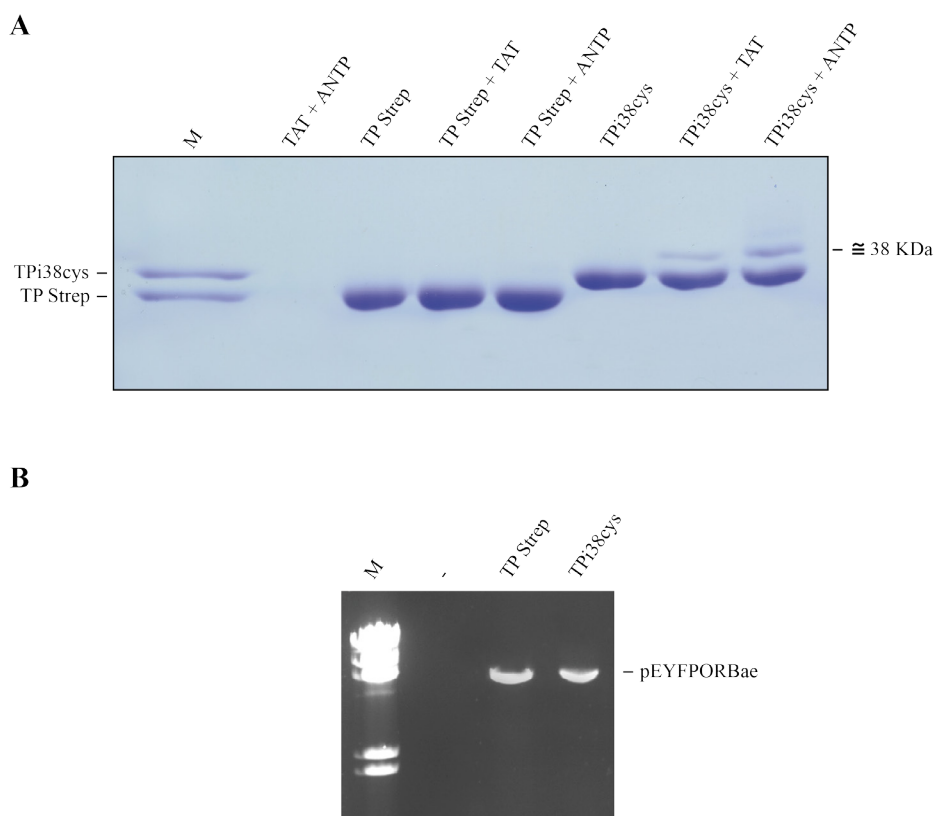
### 7.2. Análisis de la funcionalidad de la TPi38cys

En primer lugar se estudió la capacidad de conjugación que tenía la TPi38cys con los péptidos TAT y ANTP fusionados a un grupo maleimida. La reacción de conjugación se realizó en una disolución tamponada con 50 mM HEPES, pH 7.5 y se incubó toda la noche a 4°C. Se optó por el uso del tampón HEPES por ser frecuentemente utilizado en medios de cultivo para células de mamífero y en previsión de la utilización del producto de conjugación en cultivos celulares. Por otra parte, el pH óptimo para la reac- ción con la cisteína es alrededor de 7.5, mientras que con pHs superiores a 8 la reacción está favorecida para aminas primarias como la lisina y arginina. En un volumen final de 25 µl se incubó 1 µM de TPi38cys con 20 µM de péptido TAT o ANTP (20 veces la cantidad de la proteína con el grupo tiol según las reco- mendaciones de la casa comercial). El resultado de la conjugación y los respectivos controles se observa en la **Figura 38A**. Mientras que con la proteína TP Strep no se observó conjugación cuando se incubó con cualquiera de los dos péptidos, la TPi38cys sí producía un producto de movilidad electroforética distinta a consecuencia de la conjugación.

Para observar la capacidad de la proteína TPi38cys para amplificar DNA heterólogo se realiza- ron experimentos de amplificación con el plásmido pEYFPORBae linealizado. Empleando las mismas condiciones de reacción que para la proteína i38 se obtenía una cantidad muy pequeña de producto de amplificación, por lo que se realizó una puesta a punto del experimento de amplificación para la proteína



TPi38cys, resultando las nuevas condiciones experimentales las siguientes: 15 nM DNA polimerasa, 97 nM de TPi38cys, 23  $\mu$ M p5, 0.8  $\mu$ M p6 y 5 mM  $(\text{NH}_4)_2\text{SO}_4$ , con las que se consiguió un 50% del producto obtenido con la TP wt (**Figura 38B**).



**Figura 38. (A) Conjugación de la proteína TPi38cys a los péptidos TAT y ANTP por medio de un grupo maleimida.** En un volumen final de 25  $\mu$ l se incubó 1  $\mu$ M de las proteínas TP Strep y TPi38cys con 20  $\mu$ M de péptido Mal-TAT o Mal-ANTP durante toda la noche a 4°C (carriles TP Strep + TAT/TPi38cys + TAT y TP Strep + ANTP/TPi38cys + ANTP, respectivamente). Como control de la conjugación se incubó la misma cantidad de ambas proteínas en ausencia de péptidos y en las mismas condiciones (carriles TP Strep y TPi38cys). Un punto con los péptidos TAT y ANTP en ausencia de proteína y tratado en las mismas condiciones se incluyó en el experimento (carril TAT+ANTP). Como marcador de peso molecular se empleó una mezcla de las proteínas purificadas TP Strep y TPi38cys (carril M). **(B) Amplificación del plásmido pEYFPORBae con la proteína TPi38cys.** La linealización del plásmido pEYFPORBae y el experimento de amplificación se realizaron como se describe en *Materiales y Métodos* empleando 170 nM de plásmido pEYFPORBae linealizado. Las reacciones se incubaron durante 2 horas a 22°C. Como marcador de peso molecular se empleó el DNA del fago  $\lambda$  cortado con el enzima HindIII (carril M).



## Discusión



## 1. Método de amplificación de DNAs heterólogos basado en el sistema replicativo del bacteriófago $\phi 29$

Hasta la fecha no se había descrito un sistema que permitiera la amplificación de DNAs heterólogos cebado con TP. Estudios previos con el sistema replicativo del bacteriófago  $\phi 29$  habían mostrado que la iniciación en moldes de dsDNA que correspondían a los extremos del genoma de  $\phi 29$  sin una TP parental unida covalentemente a sus extremos 5' era posible. Sin embargo, la eficiencia de iniciación de estos moldes era baja, en comparación con la que se obtenía con el TP-DNA de  $\phi 29$  como molde (Méndez *et al.*, 1992), y hasta el momento no se había intentado la amplificación de moldes con los orígenes en los dos extremos de un DNA. El trabajo descrito muestra la puesta a punto de un sistema de amplificación eficiente *in vitro* empleando el sistema replicativo de  $\phi 29$  y el uso de DNAs lineales con secuencias terminales del genoma y sin TP parental unida inicialmente a los extremos 5'. En primer lugar, se desarrolló el plásmido pETORPHI, que contenía los primeros 191 y 194 pb de los extremos izquierdo y derecho del genoma de  $\phi 29$ . Como se ha comentado, la unión de estos extremos en sentido inverso genera una diana para el enzima de restricción DraI, de forma que mediante su corte, se obtiene un DNA flanqueado por la secuencia de los extremos del DNA de  $\phi 29$ . Como alternativa se diseñaron dos adaptadores mediante la hibridación de oligonucleótidos de 68 pb que correspondían a la secuencia del extremo izquierdo del DNA de  $\phi 29$ , ya que había resultado más activo que el derecho en estos ensayos. Estos adaptadores tenían en un extremo la secuencia de  $\phi 29$  y en el otro las dianas para los enzimas de restricción BsmBI y Eco0109I, respectivamente. De esta forma, estos podían ligarse a cualquier DNA con dichas dianas para ser amplificado por el sistema replicativo de  $\phi 29$ .

Dos líneas de evidencia apoyan la noción de que se está produciendo una amplificación real desde ambos extremos del DNA basada en el sistema de  $\phi 29$ : (i) la interrupción de la continuidad entre las dos secuencias de los extremos de  $\phi 29$  en la misma molécula mediante la digestión del pETORPHI con los enzimas MluI o PvuI produce una gran reducción de la cantidad de producto amplificado. Además, el bajo nivel de amplificación obtenido con la mezcla de ambas digestiones muestra que no se dan fenómenos de complementación entre fragmentos; y (ii) la cinética de amplificación con el plásmido pETORPHI es muy similar a la obtenida con el TP-DNA como molde.

La DNA polimerasa de  $\phi 29$  es muy procesiva, siendo capaz de realizar una replicación de un DNA de 70 Kb a partir de un único evento de iniciación (Blanco *et al.*, 1989), lo que la convierte en un candidato óptimo para amplificaciones de fragmentos muy largos de DNA. Esto podría utilizarse para amplificar largos fragmentos genómicos y definidos de diversos organismos.

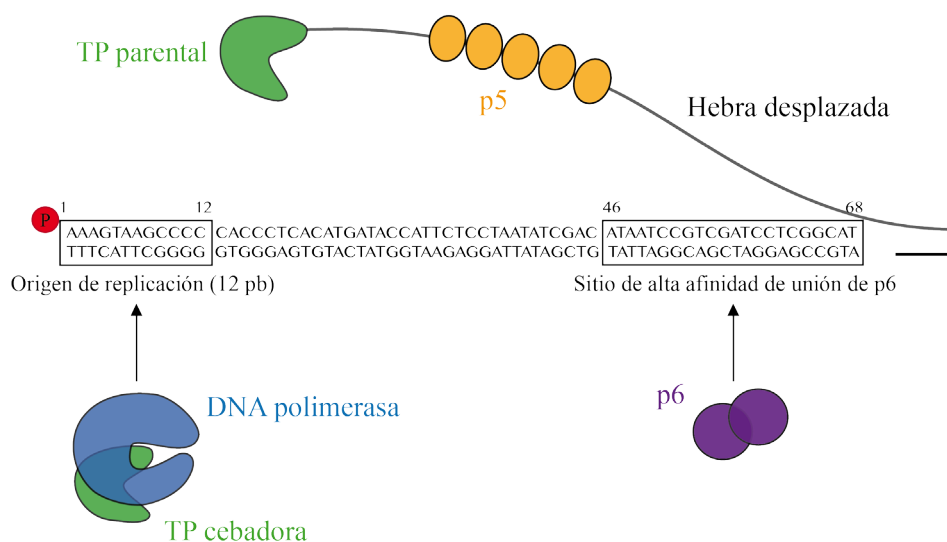
En cuanto a los requerimientos del sistema, este trabajo ha demostrado que la TP parental en el molde inicial no es esencial para la amplificación, aunque su presencia aumenta la cantidad de producto

obtenido y permite amplificar eficientemente con un 25% menos de moléculas molde. Además, la amplificación del plásmido pETORPHI es dependiente de la presencia de la proteína p6, aunque no parece ser determinante en el caso del TP-DNA en las condiciones experimentales empleadas. La **Figura 39** muestra los requerimientos mínimos del sistema en términos de proteínas y secuencias. La secuencia mínima corresponde a los 68 primeros pb del extremo izquierdo del genoma de  $\phi 29$ . Este fragmento incluía los sitios de alta afinidad para unión de p6, situado entre las posiciones 46-68. A pesar de que los sitios de unión de p6 se extienden más allá de la posición 200, aunque su afinidad a partir de esta posición es menor, nuestros resultados mostraron que el sitio de nucleación localizado entre las posiciones 46-68 de la secuencia del extremo izquierdo era suficiente para obtener amplificación estimulada por p6. Sin embargo, con los orígenes de 194 pb, el producto de amplificación obtenido era claramente superior, lo que indica que es necesario poner secuencias que aumenten la afinidad por p6. Por otra parte, estudios con moldes de DNA con y sin grupo fosfato en posición 5' de los extremos del DNA mostraron que el grupo fosfato era esencial para el éxito de la amplificación. En otros estudios no publicados se observó que dicho grupo fosfato en 5' no se puede sustituir por otros grupos funcionales como espaciadores hidrofílicos, espaciadores hidrofóbicos o la proteína biotina.

Puesto que la diana para el enzima de restricción DraI es frecuente en DNAs con un contenido medio en A y T, pensamos que podría ser un inconveniente a la hora de amplificar genomas o DNAs de muchos organismos. Por esa razón, se sustituyó por la diana para el enzima de restricción BaeI, mucho menos común, en el plásmido pETORPHIBae. El corte con este enzima produce un DNA lineal con extremos 3' protuberantes, pero se demostró que la capacidad exonucleolítica de la DNA polimerasa de  $\phi 29$  era capaz de degradar la cadena sencilla y generar extremos romos similares a los del propio genoma.

Las aplicaciones potenciales de una amplificación de DNAs heterólogos cebada por TP y basada en el sistema replicativo de  $\phi 29$  podrían ser de dos tipos: (i) Amplificación de forma delimitada DNAs largos, como genomas virales, fragmentos de cromosomas o plásmidos. Esto también elimina la necesidad de conocer la secuencia a amplificar y el diseño de oligonucleótidos específicos. La TP unida covalentemente a los extremos 5' del producto amplificado podría ser eliminada mediante el corte de los extremos por enzimas de restricción si dicha TP no es necesaria o potencialmente puede interferir en pasos posteriores. (ii) Producción de gran cantidad de DNAs unidos de forma covalente a TP de forma específica de sitio (extremos) para usos en los que una molécula híbrida de este tipo fuera requerida. Un ejemplo de ello son los procesos de transferencia génica en células de mamífero, donde la TP podría ser empleada como plataforma para la fusión de diferentes péptidos señal. Como se ha comentado en el *Capítulo 4*, el DNA podría ser marcado para ir a células específicas o compartimientos subcelulares, además de estar protegido frente a la actividad de exonucleasas en los extremos 5'. En este sentido, se determinó que la expresión *in vivo* de la proteína de fusión YFP-TP complementaba la infección de un fago mutante de  $\phi 29$  que no expresaba

la TP, indicando que la fusión de otras proteínas a la TP podría no interferir en su función (Muñoz-Espín *et al.*, 2010). Otro ejemplo podría ser la generación de moléculas híbridas enzima-aptámero que pudieran explotar la naturaleza de ambos tipos de moléculas para determinadas aplicaciones. De esta forma, un enzima o dominio con actividad podría estar fusionada a la TP, que a su vez funciona como el cebador de la amplificación de pequeños DNAs con una secuencia que se pliega en forma de aptámero. Esto supondría la generación de múltiples copias de una molécula híbrida enzima-TP/aptámero-DNA, donde el aptámero podría reclutar un molde específico para el que el componente enzimático tuviera una actividad determinada gracias a la propia unión proteína-DNA. En cuanto a la viabilidad de los DNAs amplificados con el sistema replicativo de  $\phi 29$ , trabajos previos mostraron que la funcionalidad del producto amplificado empleando el sistema mínimo de cuatro proteínas purificadas *in vitro* (DNA polimerasa, TP, p5 y p6) era idéntico al aislado de las partículas del fago y que tenía la misma capacidad de producir progenie viral en protoplastos de *B. subtilis* (Blanco *et al.*, 1994). Esto sugiere que el sistema descrito en este trabajo puede ser una herramienta potencial para la generación de TP-DNAs con actividad biológica para diversas aplicaciones.



**Figura 39. Requerimientos mínimos del origen de  $\phi 29$  para la amplificación *in vitro* de DNAs heterólogos.** En la figura se muestra el origen de 68 pb con la secuencia del extremo izquierdo del DNA de  $\phi 29$ . Los primeros 12 pb del extremo (rectángulo) se requieren para la iniciación y elongación de la cadena de nueva síntesis por parte del heterodímero (TP cebadora en verde, DNA polimerasa en azul). El grupo fosfato en el extremo 5' (P, círculo rojo) se ha observado que es esencial para la amplificación *in vitro*. Las secuencias de unión de alta afinidad (rectángulo en las posiciones 46-68) para la proteína p6 (morado) son necesarios para la estimulación de la iniciación por p6. La proteína de unión a DNA de banda simple p5 (naranja) es necesaria también para la amplificación.

### 2. Mejora de los orígenes de replicación y estudio de las primeras fases del proceso replicativo del bacteriófago $\phi 29$

El mecanismo de iniciación cebado por TP ha sido extensamente estudiado en el modelo replicativo del bacteriófago  $\phi 29$ , lo cual ha sido de gran utilidad en el estudio de sistemas con un mecanismo de replicación similar (Caldentey *et al.*, 1992; King y van der Vliet, 1994; Martín *et al.*, 1996; Illana *et al.*, 1996; González-Huici *et al.*, 1996; Longás *et al.*, 2006). Sin embargo, a pesar de las similitudes, existe un alto grado de especificidad en las interacciones entre la DNA polimerasa, la TP y los orígenes de replicación de los diferentes organismos, así como diferencias en el mecanismo de iniciación. Mientras la mayor parte de los esfuerzos se han centrado en el estudio de los mecanismos de recuperación de la información en la replicación cebada por TP, como el *sliding-back*, hasta el momento no ha habido un estudio sistemático de la secuencia de los orígenes de replicación. El análisis realizado sobre la mejora de estos orígenes en el genoma del bacteriófago  $\phi 29$  mediante la construcción de orígenes artificiales de 29-mer y 68-mer, ha proporcionado información sobre los primeros pasos del proceso replicativo iniciado por proteína terminal. Estudios previos sobre fragmentos delecionados del plásmido pID13, que contenían las secuencias de los extremos izquierdo y derecho del DNA de  $\phi 29$ , mostraron que la integridad del triplete de las ITRs del genoma era necesario para un correcto y eficiente mecanismo de *sliding-back*, aunque no se estudiaron extensivamente los efectos de secuencia (Gutiérrez *et al.*, 1988). Sí se identificaron, sin embargo, algunos cambios en las posiciones 4-12 que afectaban a la actividad del origen de replicación. Gracias al presente estudio, se observó que pueden ser construidos *in vitro* orígenes artificiales y activos mediante la combinación de diferentes bloques de 3 nucleótidos en las primeras 6 posiciones, incluso cuando las secuencias no sean complementarias y formen desapareamientos en el extremo o internos. El análisis mutacional del triplete terminal mostró que el triplete más eficiente en la hebra molde era el constituido por CCC. La iniciación con dGTP haría que la estabilidad del par G/C, mayor que el *wild-type* (A/T), estabilizara al complejo formado por el heterodímero DNA polimerasa/TP, la hebra molde y los primeros nucleótidos incorporados (Gs), produciendo un mayor número de interacciones. Esto supondría un mayor número de iniciaciones efectivas que podrían dar lugar a una mayor elongación de la cadena. Tras el triplete CCC, el molde *wild-type* (TTT) resultó el más eficiente, constituido por otra pirimidina. Por otra parte, las purinas A y G dieron como resultado una actividad muy baja, lo que podría sugerir que el complejo no acomoda bien el volumen de las purinas localizadas en la hebra molde. En los tripletes molde constituidos por T, A o G no se observaron efectos significativos con las variaciones en la secuencia del triplete de la hebra desplazada. Sin embargo, aunque se obtenían orígenes artificiales activos en todas las combinaciones en el molde con C, se observó una clara preferencia por el triplete AAA en el extremo 5'. Por otra parte, la actividad no parece depender únicamente de la naturaleza de la secuencia sino también del desaparea-



miento, puesto que se obtuvo una actividad replicativa muy baja con el origen completamente apareado G/C, posiblemente porque, en este caso, se requiere más energía para abrir la doble banda que en el origen *wild-type* o cualquier otra construcción con el extremo ya desapareado. Aunque dicho desapareamiento en el extremo parece tener un efecto general positivo, se ha de tener en cuenta la identidad de la secuencia de ambas hebras a la hora de hacer predicciones en la actividad replicativa de los orígenes sintéticos. Así, tres parámetros interrelacionados parecen influir claramente en dicha actividad: la secuencia de la hebra molde, la secuencia de la hebra desplazada y el desapareamiento del extremo.

Las posiciones 4-6 tienen una potencial variabilidad que hace que el estudio pormenorizado sea más complejo. Por esa razón, su estudio se realizó en base a una comparativa con la secuencia del bacteriófago GA-1 y la sustitución por secuencias escogidas al azar. En un contexto de apareamiento total, el efecto de la secuencia en las posiciones 4-6 no parece grande y, en la mayor parte de los casos, el cambio de la secuencia *wild-type* supuso un efecto negativo. Por otra parte, un desapareamiento interno en dichas posiciones incrementa hasta 10 veces la actividad *wild-type*. Además, dicho efecto junto con el generado por el par AAA/CCC parecía ser aditivo en la eficiencia del origen. De esa forma, la unión del efecto del desapareamiento AAA/CCC (15 veces) y el de las posiciones 4-6 (10 veces) alcanzó una actividad superior a las 25 veces en el origen 6m. El hecho de que el efecto sea aproximadamente aditivo parece suponer que ambos desapareamientos están facilitando el mismo paso en las primeras etapas de la replicación. El análisis del efecto de la secuencia en orígenes con las 6 primeras posiciones desapareadas mostró que la importancia de ésta era moderada, ya que las actividades registradas oscilaban entre un 27.6% (origen AAATCA/CCCGTA) y un 109.5% (origen AAACAT/CCCCAT) respecto a la actividad obtenida con el origen 6m. Esto podría ser explicado asumiendo que, para estos orígenes, el mayor efecto es a consecuencia del desapareamiento.

Estos resultados muestran la importancia de los primeros pasos del proceso de replicación y sugiere que las principales barreras a nivel de iniciación pueden ser superadas mediante: (i) la estabilización del complejo de iniciación, si los primeros nucleótidos incorporados dan lugar a un par de bases fuertes (G/C); (ii) la apertura de la doble hebra en el origen de replicación y (iii) la presencia AAA en el extremo 5'.

La preferencia por AAA en el extremo 5' de la hebra desplazada requiere un extremo estructuralmente reconocible, de doble banda roma, incluso cuando está parcialmente desapareado porque si añadimos una cola de ssDNA de 10 nucleótidos a la hebra desplazada no hay un efecto positivo debido a la secuencia AAA, ni cuando está en el extremo 5' protuberante ni cuando está apareada con el triplete molde. Una posible explicación sería que cuando la hebra desplazada alcanza cierta longitud, ésta cubre el surco destinado para ella en la superficie de la DNA polimerasa y evita el reconocimiento específico del extremo del DNA. Por otra parte, la preferencia por AAA en el extremo 5' del origen de replicación puede considerarse un mecanismo adicional de fidelidad porque supone la presencia de TTT en la hebra molde

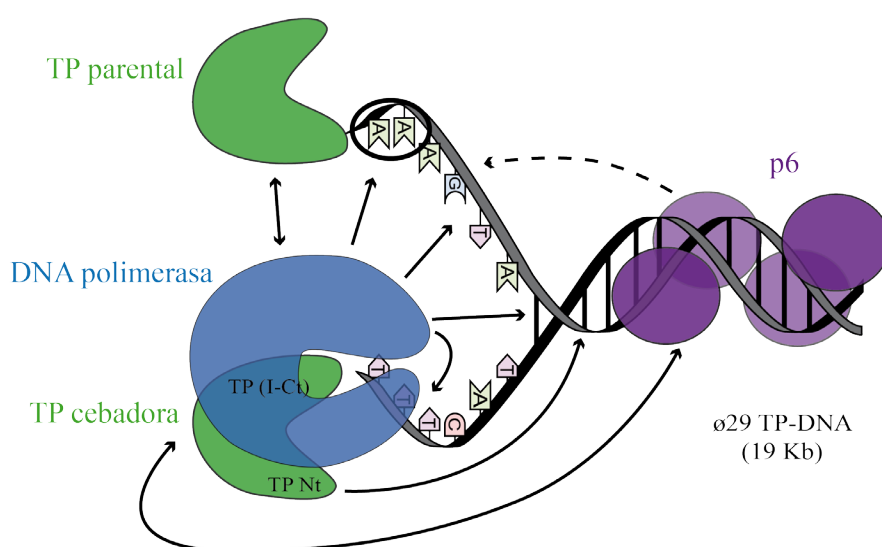
para el correcto mecanismo del *sliding back*. De esa forma, una secuencia incorrecta en el extremo molde 3' no tendría una secuencia AAA en el 5' y, por lo tanto, sería reconocida con mucha menos eficiencia por el heterodímero TP/DNA polimerasa. Otras construcciones no naturales como un extremo 5' recesivo o la presencia de una estructura tipo tallo-giro en el extremo 5' tienen una actividad no detectable, lo que subrayaría la importancia de la correcta longitud y estructura de la hebra desplazada para el mecanismo de iniciación.

Los resultados también mostraron que el dominio N-terminal de la TP, con capacidad de unión a DNA de forma independiente de secuencia, no es requerido para el reconocimiento de A en 5'. Sin embargo, sí parece crítico para una utilización eficiente de orígenes con extremos apareados e incluso con 3 desapareamientos por el heterodímero DNA polimerasa/TP, siendo más importante el efecto del dominio N-terminal cuando hay un mayor número de posiciones apareadas.

La apertura de la cadena en orígenes artificiales de 68 pb mostró que, aunque el número óptimo de posiciones desapareadas para la eficiencia del origen de replicación es de 6, una mayor apertura supone orígenes aún activos (hasta 2.4 veces la actividad del origen wt para el origen 12m). Estos datos mostraron un paralelismo con el efecto estimulador de la iniciación de la p6 observado en trabajos previos. La proteína p6 se ha propuesto como el actor principal en la apertura de la doble cadena (Prieto *et al.*, 1988; Serrano *et al.*, 1994) y, de acuerdo con ello, en nuestro análisis se describe que los orígenes con posiciones desapareadas estimulan los primeros pasos de la replicación. De hecho, la p6 produce una mayor estimulación cuanto más cerrada esté la doble cadena, lo que en la explicación más sencilla supondría que se está favoreciendo el mismo paso en el proceso replicativo. Por otra parte, la p6 ha sido propuesta como la responsable del reclutamiento de la maquinaria replicativa (heterodímero DNA polimerasa/TP) al origen de replicación mediante interacciones directas de tipo proteína-proteína (Freire *et al.*, 1996; Pérez-Arnaiz *et al.*, 2007). Puesto que un cierto nivel de estimulación por p6 se mantiene con todos los orígenes abiertos, un mecanismo adicional de estimulación de p6, como el propuesto, también es compatible con lo observado.

En cuanto al paso específico que está siendo facilitado en los orígenes A/C y 6m en la replicación cebada por TP, parece existir una clara estimulación en el proceso de iniciación, probablemente debido a la accesibilidad del heterodímero al origen facilitada por la apertura de la doble hebra y la estabilización del par G-C en el triplete iniciador. Pero sobre todo, hay una fuerte estimulación en la transición, posiblemente debido al alivio de la barrera termodinámica de la apertura de la doble cadena en el momento en que las interacciones entre la DNA polimerasa y la TP se están rompiendo (Méndez *et al.*, 1997).

Algunos de los hallazgos de este estudio muestran similitudes con los procesos replicativos de adenovirus y el fago PRD1. Por ejemplo, se ha observado que el heterodímero DNA polimerasa/pTP de adenovirus reconoce una secuencia localizada entre las posiciones 8-17 del origen, contenidas en las ITR



**Figura 40. Modelo de las interacciones durante la iniciación de la replicación de  $\phi 29$ .** Las flechas simples indican interacciones de tipo proteína-DNA, las flechas dobles indican interacciones de tipo proteína-proteína y la flecha simple con línea discontinua indica el efecto de apertura de hélice de DNA de la proteína p6. Se muestran los seis primeros nucleótidos, en los que se ha centrado el presente estudio de las bandas molde y desplazada.

Un modelo que tiene en cuenta los resultados presentados en este estudio se muestra en la **Figura 40**. El posible orden de eventos podría ser el siguiente: la TP parental y el complejo nucleoproteico p6-DNA unidos en el extremo del genoma de  $\phi 29$  reclutarían el heterodímero DNA polimerasa/TP, que se uniría a la hebra molde y reconocería el triplete AAA en el extremo 5' de la hebra desplazada. Este complejo posiblemente también interactuase con secuencias a partir de la posición 4. Al mismo tiempo, el dominio N-terminal de la TP estabilizaría el complejo mediante su unión a regiones dsDNA de forma no específica de secuencia y los nucleótidos comenzarían a ser incorporados a la hebra de nueva síntesis. La apertura de la doble cadena permitiría la unión del heterodímero

ro a la hebra desplazada, aumentando su estabilidad y la progresión a la elongación a través de la etapa de transición. En este punto, al mismo tiempo en que las interacciones entre la DNA polimerasa y la TP se están rompiendo, ha de ser superada la barrera termocinámica de la apertura de la hélice. Tras el paso a la elongación, la DNA polimerasa disociada de la TP continuaría de forma procesiva y mediante el desplazamiento de banda con la replicación del genoma completo del fago.

Los resultados presentados aquí, no solo ayudan a comprender los eventos tempranos del proceso de replicación en un origen con TP unida a su extremo 5' sino que podrían ser la base para el diseño de procesos replicativos cebados por TP con eficiencia aumentada para diversas aplicaciones. Como se ha mostrado en el epígrafe anterior, DNAs heterólogos flanqueados por los orígenes de replicación de  $\phi 29$  y que carecen de una TP covalentemente unida a los extremos 5' presentan una moderada eficiencia de amplificación, en términos de moléculas iniciales requeridas para obtener la amplificación. Por esa razón, se podrían diseñar nuevos vectores replicativos que compensaran la ausencia de TP parental, fundamentalmente en la disminución del número mínimo de moléculas molde necesarias para producir amplificación. Esto podría facilitar las aplicaciones mencionadas anteriormente en la producción eficiente de amplicones definidos unidos covalentemente a una TP.

### 3. Sustitución del dominio N-terminal de la TP de $\phi 29$ por el dominio de unión a DNA del factor de transcripción GAL4

El gran número de interacciones de tipo proteína-DNA o proteína-proteína que se producen en los orígenes de replicación del genoma de  $\phi 29$  hace que existan numerosas dianas para la mejora de la eficiencia de replicación. Puesto que el dominio N-terminal de la TP de  $\phi 29$  tiene capacidad de unión a DNA de forma inespecífica de secuencia, hemos propuesto que una forma de unión más específica podría mejorar el acceso de la maquinaria de replicación al extremo del origen de replicación. Para ello, se sustituyó el dominio N-terminal de la TP por el dominio de unión a DNA del factor de transcripción GAL4 y se generó una diana de reconocimiento para GAL4 haciendo los mínimos cambios posibles en un origen basado en la secuencia del extremo derecho del DNA de  $\phi 29$ . Experimentos con la TP *wild-type* mostraron, sorprendentemente, que el origen que contenía la secuencia UAS estaba muy afectado en su actividad de replicación, por lo que los cambios en la secuencia habían provocado una disminución en la eficiencia de utilización del origen. Dicha disminución no era predecible a la luz de los resultados obtenidos en el Capítulo 2. Por otra parte, los experimentos con la TP GAL4 mostraron que la proteína parecía inactiva en replicación tanto con el molde de DNA que tenía la secuencia *wild-type* como con el que tenía la secuencia UAS. Esto implicaba que la sustitución del dominio N-terminal de la TP por el dominio de unión de GAL4

había afectado a la funcionalidad de la TP y, como se vio por experimentos de retraso en gel, la TP GAL4 había perdido su capacidad de unión a DNA.

#### 4. Identificación de sitios potenciales de inserción de secuencias peptídicas permitidos en la TP del bacteriófago $\phi 29$

La fusión de proteínas ha sido un procedimiento ampliamente utilizado para la monitorización de proteínas o el desarrollo de plataformas donde dos actividades juntas podrían ser útiles. Tradicionalmente, esto se ha realizado mediante el clonaje de las respectivas secuencias de DNA codificantes unidas, por lo que las proteínas de fusión se suelen unir a los extremos N o C-terminales. Sin embargo, en algunas ocasiones esta fusión terminal no es funcional o bien la inserción en el medio de la secuencia supone una ventaja determinada, como la de eliminar partes de la proteína no requeridas o la disminución de la hidrofobicidad mediante la eliminación de fragmentos con gran contenido de residuos no polares. Como el hecho de encontrar sitios potenciales de inserción en el interior de una secuencia proteica supone un reto, se han desarrollado técnicas para el cribado de dichos sitios, como es el caso de la inserción al azar de un dominio peptídico.

En este sentido, la unión covalente de la TP de  $\phi 29$  al DNA la convierte en una proteína única para este tipo de abordaje, puesto que es una excelente base para añadir nuevas funciones a un DNA. Además, como se ha descrito en este mismo trabajo, el desarrollo de un sistema de amplificación de DNAs heterólogos basado en el sistema replicativo de  $\phi 29$  supone que, mediante un proceso de amplificación, cualquier DNA podría ser unido a una TP con nuevas funciones. Con el objeto de identificar los sitios permitidos de inserción en la TP de  $\phi 29$ , se realizó un cribado mediante la transposición al azar de la proteína GFP en su secuencia. La gran mayoría de los clones seleccionados que contenían la GFP en el interior de la secuencia resultaban inactivos en experimentos de iniciación de la replicación con extractos bacterianos empleando el TP-DNA de  $\phi 29$  como molde, por lo que se redujo el tamaño de la GFP a una inserción de 15 aminoácidos. De esta forma se purificaron las TPs con inserciones en las siguientes posiciones: i38, i65, i70, i75, i115, i118, i144 e i174. Además, se purificaron dos inserciones con la GFP completa: GFP14 y GFP38. Por otra parte, se construyeron mediante la fusión de la YFP o GFP completa en los extremos N y C terminales las proteínas YFP TP y TP GFP, respectivamente.

Las fusiones a los extremos N y C terminales (YFP TP y TP GFP, respectivamente) resultaron en proteínas activas en iniciación de la replicación y en amplificación, lo que indicaba que fusiones en los extremos de la TP eran permitidas. Además, se ha demostrado que la YFP TP es activa en experimentos *in vivo* (Redrejo *et al.*, 2013) y su estudio en la presente tesis mediante experimentos de competitividad

mostró que tenía un comportamiento muy similar al de su correspondiente control *wild-type*, que tiene fusionado al extremo N-terminal un doble Strep-tag (TP Strep). Por otra parte, la amplificación con la TP GFP del plásmido pETORPHI linealizado no daba como resultado una única banda correspondiente al tamaño unidad. Experimentos de competitividad mostraron que ésta apenas competía a la TP *wild-type* (TP Strep) en su interacción con la DNA polimerasa/TP-DNA, posiblemente debido a que el volumen de la GFP en el extremo C-terminal de la TP supone un problema en la acomodación de la TP en la DNA polimerasa para formar el heterodímero, haciéndolo más inestable y/o interfiriendo con la funcionalidad del bucle iniciador. La presencia de TP parental en el DNA molde permitiría su estabilización mediante el incremento del número de interacciones, resultando en una amplificación más eficiente del TP-DNA que en ausencia de dicha TP parental.

Las proteínas con inserciones en, o próximas, al dominio intermedio de la TP son proteínas que conservan la capacidad de interacción con DNA pero son muy poco activas. Estos resultados son explicables ya que, como se ha explicado en la Introducción, el dominio intermedio de la TP de  $\phi 29$  es el principal responsable de la estabilidad de la formación del heterodímero mediante la formación de interacciones con el subdominio TPR1 de la DNA polimerasa (Pérez-Arnáiz *et al.*, 2007). Las proteínas i144 e i174 tienen una actividad muy baja en iniciación y son inactivas en amplificación. Además, el producto de iniciación correspondiente a estas TP no se detectó en los experimentos de competitividad. En este caso, la inserción de 15 aminoácidos podría estar afectando a la interacción con la DNA polimerasa ya que dos residuos próximos a estas posiciones como son el R158 y el R169, respectivamente, han sido descritos como fundamentales en la estabilidad del complejo DNA polimerasa/TP y están conservados en otros fagos como GA-1 (Kamtekar *et al.*, 2006; del Prado *et al.*, 2012). Los residuos R158 y R169 de la TP interaccionan de forma específica con los residuos E291 y E322 de la DNA polimerasa, respectivamente, por lo que la inserción de una secuencia peptídica en sus proximidades podría provocar que ambos residuos no pudieran contactar con sus pares respectivos, impidiendo la formación correcta del heterodímero. El alineamiento de la secuencia de aminoácidos de la TP del bacteriófago MG-B1 con la de  $\phi 29$  mostró inserciones peptídicas en las posiciones 142 y 158, que apoyaban el hecho observado de que ciertas inserciones en la secuencia de la TP de  $\phi 29$  pueden ser permitidas. La inserción en la posición 142 está muy próxima a la inserción 144 descrita en este trabajo, lo que sugiere que nuestro proceso de selección ha sido efectivo, detectando los posibles sitios de inserción funcionales. Sin embargo, como se ha comentado anteriormente en el caso de  $\phi 29$ , inserciones en el dominio intermedio y, concretamente, en las posiciones 144 y 174 resultaban en proteínas casi inactivas, por lo que las inserciones observadas en el fago MG-B1 están evolutivamente adaptadas a la función de la proteína.

Las proteínas i115 e i118 tienen un comportamiento muy similar entre sí, con un nivel de iniciación de la replicación *wild-type*, pero inactivas en amplificación de DNA. Sin embargo, los experimentos de

competitividad mostraron altos niveles de actividad para ambas, siendo la proteína i115 la única de las TPs derivadas de este estudio que interaccionaba con la DNA polimerasa/TP-DNA mejor que la propia TP *wild-type*. La interacción entre el dominio intermedio de la TP y el TPR1 de la DNA polimerasa es el resultado, en su mayor parte, de interacciones de tipo hidrofóbico. El análisis del valor GRAVY del péptido de 15 aminoácidos en la escala de hidropatía desarrollada por Kyte y Doolittle (1982) dio como resultado -0.176. Mientras que valores positivos del valor GRAVY indican que el péptido es polar, valores negativos indican que el péptido es no polar. Posiblemente, la inserción de un péptido hidrofóbico en la posición 115 incrementa el número de interacciones con la DNA polimerasa, produciendo una alta estabilización del heterodímero DNA polimerasa/TP que incluso parece impedir el progreso con la elongación de la cadena de nueva síntesis. Este hallazgo podría suponer la base para el desarrollo de un inhibidor potencial de la DNA polimerasa. Por otra parte, la proteína i75 se vio afectada en menor medida puesto que, a pesar de tener una baja actividad de iniciación de replicación, mantenía aún un 50% de la actividad en la amplificación con TP-DNA como molde. Posiblemente, el efecto no sea tan drástico como en los casos anteriores al encontrarse dicha posición en un bucle entre los dominios N-terminal e intermedio de la TP de  $\phi 29$ .

Como se ha mencionado anteriormente, el dominio N-terminal presenta capacidad de unión a DNA de forma inespecífica de secuencia (Zaballos y Salas, 1989) que la hace localizar en el nucleóide bacteriano (Muñoz-Espín *et al.*, 2010). Además, se ha descrito que el fragmento mínimo de la TP para la unión a DNA consiste en los 70 primeros aminoácidos (Holguera *et al.*, 2014). Posiblemente por esa razón, inserciones en el medio de dicha secuencia, como en el caso de las proteínas i65 e i70, mostraron problemas en la unión a DNA, ya que a la mayor concentración de proteína analizada ( $[TP] = 40 \text{ nM}$ ) en ensayos de retraso en gel aún había DNA libre. Así, la inestabilidad en el origen de replicación a consecuencia de un defecto en la interacción del heterodímero con el DNA, podría ser la consecuencia de la baja actividad de iniciación observada con estas proteínas. Por otra parte, cuando la TP está completa, se ha descrito que los primeros 15 aminoácidos son prescindibles para la unión a DNA (Holguera *et al.*, 2014), lo que supone que cierta capacidad de interacción con el DNA se mantiene en el resto de la proteína. Probablemente por esta razón, la inserción de una GFP completa en la posición 14, GFP14, no parece afectar a la actividad de la TP, que muestra una actividad de iniciación *wild-type* y una actividad de amplificación con TP-DNA cercana al 80% la de la *wild-type*. Esto tiene especial relevancia cuando se compara con la proteína GFP38, que contiene la misma inserción más allá de esos 15 aminoácidos, ya que esta proteína es inactiva en todas las actividades ensayadas. Por su parte, cabe destacar que cuando dicha GFP en posición 38 se reduce a la inserción peptídica de 15 aminoácidos (i38), la actividad se recupera, siendo activa en iniciación y en amplificación tanto con TP-DNA como con el pETORPHI linealizado. Esto indica que la posición 38 permite una inserción de 15 aminoácidos, pero no una del tamaño de la GFP.

La proteína i38 se muestra como una potencial herramienta biotecnológica por dos razones princi-



pales: (i) permite la amplificación de DNAs heterólogos que no tienen una TP unida covalentemente a su extremo 5' y (ii) mediante la adición de una diana para una proteasa en dicha posición, se pueden eliminar las actividades de unión a DNA de forma inespecífica de secuencia y la señal de localización nuclear eucariótica (NLS) (Redrejo-Rodríguez *et al.*, 2012; Redrejo-Rodríguez y Salas, 2014) contenidas hacia el extremo N-terminal, en caso de no ser requeridas. Cabe recordar que el dominio N-terminal es parcialmente dispensable para la replicación *in vitro* del DNA viral en presencia de p6 (Pérez-Arnáiz *et al.*, 2007) y que la eliminación de los primeros 38 aminoácidos reduciría el tamaño de la TP aproximadamente en un 15%. Para probar el potencial biotecnológico de la proteína i38 se construyó la proteína TPi38cys, en la que se había sustituido la inserción de 15 aminoácidos por un fragmento peptídico que contenía una diana para la trombina en la parte N-terminal y una cisteína a modo de grupo reactivo que permitiera la adición de dominios proteicos por medio de la reacción con el grupo tiol del residuo aminoacídico. La conjugación se puso a punto con los péptidos de penetración celular (CPPs) TAT y ANTP (antennapedia), y como grupo reactivo se empleó la maleimida, que reacciona de forma espontánea con el grupo tiol de la cisteína. El análisis proteómico mostró que, efectivamente, se producía la conjugación específica de dichos péptidos.

La posibilidad de fusionar péptidos en *trans* a una TP que se une covalentemente a su ácido nucleico codificante hace del sistema una plataforma muy interesante para diversas aplicaciones en las que es útil la unión física proteína-DNA. Por una parte, la proteína TPi38cys podría ser la base para la optimización de procesos de transferencia génica de forma pasiva, pudiendo incluso especificar el destino del DNA al que está covalentemente unida mediante la conjugación con péptidos que incluyan señales de localización subcelular. Además, como se ha mencionado anteriormente, la TPi38cys podría ser un buen candidato para procesos de cribado de librerías polipeptídicas en busca de propiedades específicas o como complemento a la tecnología aptamérica. Por otra parte, la posibilidad de conjugar péptidos o dominios peptídicos con actividades específicas, hace de la TPi38cys un potencial soporte para la adición de funciones en serie de forma modular, pudiendo llegarla a convertir en un microsistema de gran versatilidad. En esta línea, la TPi38cys podría incorporar, además de las señales de localización subcelular, sistemas de integración génica como recombinasas o transposasas que permitiera integrar el DNA al que está unido covalentemente al genoma de las células o al de compartimentos subcelulares como el mitocondrial.



## Conclusiones



1. El uso como molde de DNAs lineales flanqueados por las secuencias terminales del genoma del bacteriófago  $\phi 29$  y sin TP parental unida inicialmente a los extremos 5' permite su amplificación eficiente *in vitro* empleando el sistema replicativo de  $\phi 29$ : DNA polimerasa, TP, p5 y p6.
2. Se han desarrollado dos formas de producir DNAs heterólogos flanqueados por las secuencias terminales del genoma de  $\phi 29$ : (i) plásmidos que mediante un corte con el enzima de restricción DraI o BaeI producen un DNA flanqueado por los primeros 191 y 194 pares de bases con la secuencia de los extremos izquierdo y derecho, respectivamente, del genoma de  $\phi 29$ ; (ii) *linkers* de dsDNA de 68 pares de bases con la secuencia del extremo izquierdo del genoma de  $\phi 29$  pueden ser ligados a cualquier DNA.
3. El grupo fosfato en el extremo 5' de la hebra desplazada es indispensable para la amplificación de DNAs heterólogos flanqueados por las secuencias terminales del genoma de  $\phi 29$ .
4. Tres parámetros parecen influir en la eficiencia de utilización de los orígenes de replicación: la secuencia de la hebra molde, la secuencia de la hebra desplazada y el desapareamiento en el extremo.
5. El mayor efecto en la eficiencia de los orígenes artificiales de dsDNA se logra mediante un desapareamiento de 6 pares de bases en el extremo, aunque una mayor apertura de la doble cadena mantiene un cierto grado de estimulación (2.4 veces la actividad *wild-type* para un origen con 12 posiciones desapareadas). El desapareamiento de los orígenes parece estar favoreciendo el mismo paso que la proteína p6.
6. Existe una preferencia por el triplete AAA en el extremo 5' de la hebra desplazada, posiblemente como un mecanismo adicional de fidelidad. En el reconocimiento de dicho triplete no participa el dominio N-terminal de la TP. En el contexto de este estudio, en la hebra molde el triplete óptimo es CCC.
7. La sustitución del dominio N-terminal de la TP de  $\phi 29$  por el dominio de unión a DNA del factor de transcripción GAL4 resultó en una TP que carecía de capacidad de unión a DNA.
8. La transposición al azar de GFP mediante el transposón Tn5 ha posibilitado la identificación de la posición 14 como sitio permisivo para la inserción funcional de GFP. Por otra parte, la sustitución de la GFP por una inserción de 15 aminoácidos ha posibilitado la identificación de las posiciones 38, 65, 70 y 75 como sitios permisivos que dan lugar a TPs activas en la amplificación de TP-DNA. Sin embargo, únicamente la inserción en la posición 38 fue activa en la amplificación del DNA lineal sin TP parental pETORPHI.
9. La inserción de 17 aminoácidos en la posición 38 de la TP de  $\phi 29$  que contenía una cisteína dio lugar a una TP activa en amplificación del DNA lineal sin TP parental pEYFPORBae. Se ha intentado la conjugación en *trans* de péptidos a dicha TP mediante la reacción con un grupo maleimida con los péptidos de penetración celular (CPPs) TAT y Antennapedia, pero los análisis proteómicos no permitieron verificar el producto obtenido como conjugado.







- Ackermann, H.W. (1998) Tailed bacteriophages: the order *Caudovirales*. *Adv Virus Res*, **51**, 135-201.
- Ahmed, S.A., Sandai, D.A., Musa, S., Hoe, C.H., Riadzi, M., Lau, K.L., y Tang, T.H. (2012) Rapid diagnosis of leptospirosis by multiplex PCR. *Malays J Med Sci*, **19**, 9.
- Alcorlo, M., González-Huici, V., Hermoso, J.M., Meijer, W.J. y Salas, M. (2007) The phage  $\phi 29$  membrane protein p16.7, involved in DNA replication, is required for efficient ejection of the viral genome. *J Bacteriol*, **189**, 5542-5549.
- Arber, W. (1965) Host-controlled modification of bacteriophage. *Annu Rev Microbiol*, **19**, 365-378.
- Asensio, J.L., Pérez-Lago, L., Lázaro, J.M., Gonzalez, C., Serrano-Heras, G., y Salas, M. (2011) Novel dimeric structure of phage  $\phi 29$ -encoded protein p56: insights into uracil-DNA glycosylase inhibition. *Nucleic Acids Res*, **39**, 9779-9788.
- Avery, O.T., MacLeod, C.M. y McCarty, M. (1944) Studies on the chemical nature of the substance introducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med*, **79**, 137-158.
- Ballesteros-Plaza, D., Holguera, I., Scheffers, D.J., Salas, M. y Muñoz-Espín, D. (2013) Phage  $\phi 29$  protein p1 promotes replication by associating with the FtsZ ring of the divisome in *Bacillus subtilis*. *Proc Natl Acad Sci U S A*, **110**, 12313-12318.
- Bamford, D.H., Caldentey, J. y Bamford, J.K. (1995) Bacteriophage PRD1: A broad host range dsDNA tectivirus with an internal membrane. *Adv Virus Res*, **45**, 281-319.
- Bao, K. y Cohen, S.N. (2003) Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *Genes Dev*, **17**, 774-785.
- Barany, F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc Natl Acad Sci U S A*, **88**, 189-193.

- Beadle, G.W. y Tatum, E.L. (1941) Genetic control of biochemical reactions in *Neurospora*. *Proc Natl Acad Sci U S A*, **27**, 499.
- Berman, A.J., Kamtekar, S., Goodman, J.L., Lázaro, J.M., de Vega, M., Blanco, L., Salas, M. y Steitz, T.A. (2007) Structures of  $\phi$ 29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J*, **26**, 3494-3505.
- Bernad, A., Zaballo, A., Salas, M. y Blanco, L. (1987) Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases. *EMBO J*, **6**, 4219-4225.
- Bhasin, A., Goryshin, I.Y., y Reznikoff, W.S. (1999) Hairpin formation in Tn5 transposition. *J Biol Chem*, **274**, 37021-37029.
- Billerbeck, S., Calles, B., Müller, C.L., Lorenzo, V. y Panke, S. (2013) Towards functional orthogonalization of protein complexes: individualization of GroEL monomers leads to distinct quasi-homogeneous single rings. *Chembiochem*. **14**, 2310-2321.
- Biondi, R.M., Baehler, P.J., Reymond, C.D. y Veron, M. (1998) Random insertion of GFP into the cAMP-dependent protein kinase regulatory subunit from *Dictyostelium discoideum*. *Nucleic Acids Res.* **26**, 4946-4952.
- Blanco, L. y Salas, M. (1984) Characterization and purification of a phage  $\phi$ 29-encoded DNA polymerase required for the initiation of replication. *Proc Natl Acad Sci U S A*, **81**, 5325-5329.
- Blanco, L., García, J.A., y Salas, M. (1984) Cloning and expression of gene 2, required for the protein-primed initiation of the Bacillus subtilis phage phi 29 DNA replication. *Gene*, **29**, 33-40.
- Blanco, L. y Salas, M. (1985) Replication of phage  $\phi$ 29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length  $\phi$ 29 DNA. *Proc Natl Acad Sci U S A*, **82**, 6404-6408.
- Blanco, L., y Salas, M. (1985). Characterization of a 3'-5' exonuclease activity in the phage  $\phi$ 29-encoded DNA polymerase. *Nucleic Acids Res*, **13**, 1239-1249.



- Blanco, L., Gutiérrez, J., Lázaro, J.M., Bernad, A. y Salas, M. (1986) Replication of phage  $\phi$ 29 DNA *in vitro*: role of the viral protein p6 in initiation and elongation. *Nucleic Acids Res*, **14**, 4923-4937.
- Blanco, L., Bernad, A., Lázaro, J.M., Martin, G., Garmendia, C. y Salas, M. (1989) Highly efficient DNA synthesis by the phage  $\phi$ 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem*, **264**, 8935-8940.
- Blanco, L., Lázaro, J.M., de Vega, M., Bonin, A. y Salas, M. (1994) Terminal protein-primed DNA amplification. *Proc Natl Acad Sci U S A*, **91**, 12198-12202.
- Blanco, L., y Salas, M. (1995) Mutational analysis of bacteriophage  $\phi$ 29 DNA polymerase. *Methods Enzymol*, **262**, 283-294.
- Blanco, L., y Salas, M. (1996). Relating structure to function in  $\phi$ 29 DNA polymerase. *J Biol Chem*, **271**, 8509-8512.
- Blasco, M.A., Blanco, L., Pares, E., Salas, M., y Bernad, A. (1990) Structural and functional analysis of temperature-sensitive mutants of the phage  $\phi$ 29 DNA polymerase. *Nucleic Acids Res*, **18**, 4763-4770.
- Blasco, M.A., Bernad, A., Blanco, L., y Salas, M. (1991) Characterization and mapping of the pyrophosphorolytic activity of the phage  $\phi$ 29 DNA polymerase. Involvement of amino acid motifs highly conserved in alpha-like DNA polymerases. *J Biol Chem*, **266**, 7904-7909.
- Bravo, A. y Salas, M. (1997) Initiation of bacteriophage  $\phi$ 29 DNA replication *in vivo*: assembly of a membrane-associated multiprotein complex. *J Mol Biol*, **269**, 102-112.
- Caldentey, J., Blanco, L., Savilahti, H., Bamford, D.H. y Salas, M. (1992) *In vitro* replication of bacteriophage PRD1 DNA. Metal activation of protein-primed initiation and DNA elongation. *Nucleic Acids Res*, **20**, 3971-3976.
- Cha, R.S. y Thilly, W.G. (1995) Specificity, efficiency, and fidelity of PCR. En Diffenbach, C.W. y Dveksler, G.S. (Eds): Specificity, Efficiency, and Fidelity of PCR in PCR Primer-A Laboratory Manual. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 37-51.

- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. y Prasher, D.C. (1994) Green fluorescent protein as a marker for gene expression. *Science*, **263**, 802-805.
- Chalfie, M. (1995) Green fluorescent protein. *Photochem Photobiol.* **62**, 651-656.
- Chamberlain, J.S., Gibbs, R.A., Rainer, J.E., Nguyen, P.N., y Thomas, C. (1988) Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res*, **16**, 11141-11156.
- Chang, J.M., Di Tommaso, P., y Notredame, C. (2014) TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol Biol Evol*, **31**, 1625-1637.
- Chang, P.C., Kim, E.S. y Cohen, S.N. (1996) *Streptomyces* linear plasmids that contain a phage-like, centrally located, replication origin. *Mol Microbiol*, **22**, 789-800.
- Chang, P.C., y Cohen, S.N. (1994) Bidirectional replication from an internal origin in a linear *Streptomyces* plasmid. *Science*, **265**, 952-954.
- Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, **6**, 201-209.
- Collins, M.L., Irvine, B., Tyner, D., Fine, E., Zayati, C., Chang, C. A., Horn, T., Ahle, D., Detmer, L.S., Kolberg, J., Bushnell, S. Urdea, M.S. y Ho, D.D. (1997) A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res*, **25**, 2979-2984.
- Compton, J. (1991) Nucleic acid sequence-based amplification. *Nature*, **350**, 91-92.
- d'Herelle, F. (1917) An invisible microbe that is antagonistic to the dysentery bacillus. *Comptes rendus Acad Sci Paris*, **165**, 373-375.
- Danna, K., y Nathans, D. (1971) Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc Natl Acad Sci U S A*, **68**, 2913-2917.

- de Jong, R.N., van der Vliet, P.C. y Brenkman, A.B. (2003) Adenovirus DNA replication: protein priming, jumping back and the role of the DNA binding protein DBP. *Curr Top Microbiol Immunol*, **272**, 187-211.
- de Vega, M., Lázaro, J.M., Mencía, M., Blanco, L., y Salas, M. (2010) Improvement of  $\phi$ 29 DNA polymerase amplification performance by fusion of DNA binding motifs. *Proc Natl Acad Sci U S A*, **107**, 16506-16511.
- Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M. y Lasken, R.S. (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A*, **99**, 5261-5266.
- Dean, F.B., Nelson, J.R., Giesler, T.L., y Lasken, R. S. (2001) Rapid amplification of plasmid and phage DNA using  $\phi$ 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res*, **11**, 1095-1099.
- del Prado, A., Lázaro, J.M., Longás, E., Villar, L., de Vega, M., y Salas, M. (2015) Insights into the determination of the templating nucleotide at the initiation of  $\phi$ 29 DNA Replication. *J Biol Chem*, **290**, 27138-27145.
- del Prado, A., Villar, L., de Vega, M. y Salas, M. (2012) Involvement of residues of the  $\phi$ 29 terminal protein intermediate and priming domains in the formation of a stable and functional heterodimer with the replicative DNA polymerase. *Nucleic Acids Res*, **40**, 3886-3897.
- Dufour, E., Méndez, J., Lázaro, J.M., de Vega, M., Blanco, L., y Salas, M. (2000) An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. *J Mol Biol*, **304**, 289-300.
- Dunning, A.M., Talmud, P. y Humphries, S.E. (1988) Errors in the polymerase chain reaction. *Nucleic Acids Res*, **16**, 10393.
- Endter, C., y Dobner, T. (2004) Cell transformation by human adenoviruses. *Curr Top Microbiol Immunol*, **273**, 163-214.

- Ereky, K. (1919) Biotechnologie der fleisch-, fett-und milcherzeugung im landwirtschaftlichen grossbetriebe für naturwissenschaftlich gebildete landwirte, verfasst. Berlin: P. Parey.
- Escarmís, C., y Salas, M. (1981) Nucleotide sequence at the termini of the DNA of *Bacillus subtilis* phage  $\phi$ 29. *Proc Natl Acad Sci U S A*, **78**, 1446-1450.
- Escarmís, C., Guirao, D., y Salas, M. (1989) Replication of recombinant  $\phi$ 29 DNA molecules in *Bacillus subtilis* protoplasts. *Virology*, **169**, 152-160.
- Esteban, J.A., Salas, M. y Blanco, L. (1993) Fidelity of  $\phi$ 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J Biol Chem*, **268**, 2719-2726.
- Esteban J.A., Blanco L., Villar L. y Salas M. (1997) *In vitro* evolution of terminal proteincontaining genomes. *Proc Natl Acad Sci U S A*, **94**, 2921-2926.
- Franklin, R. y Gosling, R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740-741.
- Freire, R., Serrano, M., Salas, M. y Hermoso, J.M. (1996) Activation of replication origins in  $\phi$ 29- related phages requires the recognition of initiation proteins to specific nucleoprotein complexes. *J Biol Chem*, **271**, 31000-31007.
- Fricova, D., Valach, M., Farkas, Z., Pfeiffer, I., Kucsera, J., Tomaska, L., y Nosek, J. (2010) The mitochondrial genome of the pathogenic yeast *Candida subhashii*: GC-rich linear DNA with a protein covalently attached to the 5' termini. *Microbiology*, **156**, 2153-2163.
- García, J.A., Peñalva, M.A., Blanco, L. y Salas, M. (1984) Template requirements for initiation of phage  $\phi$ 29 DNA replication *in vitro*. *Proc Natl Acad Sci U S A*, **81**, 80-84.
- García, P., Martín, A.C. y López, R. (1997) Bacteriophages of *Streptococcus pneumoniae*: A molecular approach. *Microb Drug Resist*, **3**, 165-176.
- Garmendia, C., Hermoso, J.M. y Salas, M. (1990) Functional domain for priming activity in the phage  $\phi$ 29 terminal protein. *Gene*, **88**, 73-79.
- Garver, K. y Guo, P. (1997) Boundary of pRNA functional domains and minimum pRNA sequence requirement for specific connector binding and DNA packaging of phage  $\phi$ 29. *Rna*, **3**, 1068.

- González-Huici, V., Lázaro, J.M., Salas, M. y Hermoso, J.M. (2000a) Specific recognition of parental terminal protein by DNA polymerase for initiation of protein-primed DNA replication. *J Biol Chem*, **275**, 14678-14683.
- González-Huici, V., Salas, M. y Hermoso J.M. (2000b) Sequence requirements for protein-primed initiation and elongation of phage  $\phi$ 29 DNA replication. *J Biol Chem*, **275**, 40547-40553.
- González-Huici, V., Salas, M. and Hermoso, J.M. (2004a) The push-pull mechanism of bacteriophage  $\phi$ 29 DNA injection. *Mol Microbiol*, **52**, 529-540.
- González-Huici, V., Alcorlo, M., Salas, M. y Hermoso, J.M. (2004b) Binding of phage  $\phi$ 29 architectural protein p6 to the viral genome: evidence for topological restriction of the phage linear DNA. *Nucleic Acids Res*, **32**, 3493-3502.
- Guatelli, J.C., Whitfield, K.M., Kwoh, D.Y., Barringer, K.J., Richman, D.D., y Gingeras, T.R. (1990) Isothermal *in vitro* amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proc Natl Acad Sci U S A*, **87**, 1874-1878.
- Gutiérrez, C., Martín, G., Sogo, J.M. y Salas, M. (1991) Mechanism of stimulation of DNA replication by bacteriophage  $\phi$ 29 single-stranded DNA-binding protein p5. *J Biol Chem*, **266**, 2104-2111.
- Gutiérrez, J., Vinós, J., Prieto, I., Méndez, E., Hermoso, J.M., y Salas, M. (1986) Signals in the  $\phi$ 29 DNA-terminal protein template for the initiation of phage  $\phi$ 29 DNA replication. *Virology*, **155**, 474-483.
- Gutiérrez, J., Garmendia, C. y Salas, M. (1988) Characterization of the origins of replication of bacteriophage  $\phi$ 29 DNA. *Nucleic Acids Res*, **16**, 5895-5914.
- Harding, N.E. y Ito, J. (1980) DNA replication of bacteriophage  $\phi$ 29: characterization of the intermediates and location of the termini of replication. *Virology*, **104**, 323-338.
- Hermoso, J.M., Méndez, E., Soriano, F. y Salas, M. (1985) Location of the serine residue involved in the linkage between the terminal protein and the DNA of phage  $\phi$ 29. *Nucleic Acids Res*, **13**, 7715-7728.

- Hinnebusch, J. y Tilly, K. (1993) Linear plasmids and chromosomes in bacteria. *Mol. Microbiol.*, **10**, 917-922.
- Hoeprich, S., Zhou, Q., Guo, S., Shu, D., Qi, G., Wang, Y., y Guo, P. (2003) Bacterial virus  $\phi$ 29 pRNA as a hammerhead ribozyme escort to destroy hepatitis B virus. *Gene therapy*, **10**, 1258-1267.
- Holguera, I., Muñoz-Espín, D., y Salas, M. (2015) Dissecting the role of the  $\phi$ 29 terminal protein DNA binding residues in viral DNA replication. *Nucleic acids Res*, **43**, 2790-2801.
- Holguera, I., Ballesteros-Plaza, D., Muñoz-Espín, D. y Salas, M. (2012) Disclosing the in vivo organization of a viral histone-like protein in *Bacillus subtilis* mediated by its capacity to recognize the viral genome. *Proc Natl Acad Sci U S A*, **109**, 5723-5728.
- Holguera, I., Redrejo-Rodríguez, M., Salas, M. y Muñoz-Espín, D. (2014) New insights in the  $\phi$ 29 terminal protein DNA-binding and host nucleoid localization functions. *Mol Microbiol*, **91**, 232-241.
- Huang, C.H., Tsai, H.H., Tsay, Y.G., Chien, Y.N., Wang, S.L., Cheng, M.Y., Ke C.H. y Chen, C.W. (2007) The telomere system of the *Streptomyces* linear plasmid SCP1 represents a novel class. *Mol Microbiol*, **63**, 1710-1718.
- Hughes, T.E., Zhang, H., Logothetis, D.E. y Berlot C.H. (2001) Visualization of a functional  $G\alpha_q$ -green fluorescent protein fusion in living cells. Association with the plasma membrane is disrupted by mutational activation and by elimination of palmitoylation sites, but not by activation mediated by receptors or AlF<sub>4</sub>. *J Biol Chem*. **276**, 4227-4235.
- Illana, B., Blanco, L. y Salas, M. (1996) Functional characterization of the genes coding for the terminal protein and DNA polymerase from bacteriophage GA-1. Evidence for a sliding-back mechanism during protein-primed GA-1 DNA replication. *J Mol Biol*, **264**, 453-464.
- Illana, B., Lázaro, J.M., Gutiérrez, C., Meijer, W.J., Blanco, L. y Salas, M. (1999) Phage  $\phi$ 29 terminal protein residues Asn80 and Tyr82 are recognition elements of the replication origins. *J Biol Chem*, **274**, 15073-15079.

- Inciarte, M.R., Salas, M. y Sogo, J.M. (1980) Structure of replicating DNA molecules of *Bacillus subtilis* bacteriophage  $\phi 29$ . *J Virol*, **34**, 187-199.
- Kamtekar, S., Berman, A.J., Wang, J., Lázaro, J.M., de Vega, M., Blanco, L., Salas, M. and Steitz, T.A. (2004) Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage  $\phi 29$ . *Mol Cell*, **16**, 609-618.
- Kamtekar, S., Berman, A.J., Wang, J., Lázaro, J.M., de Vega, M., Blanco, L., Salas, M. y Steitz, T.A. (2006) The  $\phi 29$  DNA polymerase:protein-primer structure suggests a model for the initiation to elongation transition. *EMBO J*, **25**, 1335-1343.
- Kapitonov, V.V. y Jurka, J. (2006) Self-synthesizing DNA transposons in eukaryotes. *Proc Natl Acad Sci U S A*, **103**, 4540-4545.
- King, A.J. y van der Vliet, P.C. (1994) A precursor terminal protein-trinucleotide intermediate during initiation of adenovirus DNA replication: regeneration of molecular ends in vitro by a jumping back mechanism. *EMBO J*, **13**, 5786-5792.
- Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I. y Khorana, H.G. (1971) Studies on polynucleotides: XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. *J Mol Biol*, **56**, 341-346.
- Krishna, M.M. y Englander, S.W. (2005) The N-terminal to C-terminal motif in protein folding and function. *Proc Natl Acad Sci U S A*, **102**, 1053-1058.
- Kyte, J. y Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, **157**, 105-132.
- Lawyer, F., Stoffer, S., Saiki, R., Chang, S., Landre, P., Abramson, R. y Gelfand, D. (1993) High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods Appl*, **2**, 275-287.
- Lázaro J.M., Blanco L. y Salas, M. (1995) Purification of bacteriophage  $\phi 29$  DNA polymerase. *Methods Enzymol*, **262**, 42-49.

- Liu, H., Naismith, J.H. y Hay, R.T. (2003) Adenovirus DNA replication. *Curr Top Microbiol Immunol*, **272**, 131-164.
- Lizardi, P.M., Guerra, C.E., Lomeli, H., Tusaie-Luna, I. y Kramer, R. (1988) Exponential amplification of recombinant-RNA hybridization probes. *Nature Biotechnology*, **6**, 1197-1202.
- Longás, E., de Vega, M., Lázaro, J.M. y Salas, M. (2006) Functional characterization of highly processive protein-primed DNA polymerases from phages Nf and GA-1, endowed with a potent strand displacement capacity. *Nucleic Acids Res*, **34**, 6051-6063.
- Longás, E., Villar, L., Lázaro, J.M., de Vega, M. y Salas, M. (2008) Phage  $\phi$ 29 and Nf terminal protein-priming domain specifies the internal template nucleotide to initiate DNA replication. *Proc Natl Acad Sci USA*, **105**, 18290-18295.
- Maniatis, T., Fritsch, E.F., y Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Marshall, J., Molloy, R., Moss, G.W., Howe, J.R. y Hughes, T.E. (1995) The jellyfish green fluorescent protein: a new tool for studying ion channel expression and function. *Neuron*. **14**, 211-215.
- Martín, A.C., Blanco, L., García, P., Salas, M. y Méndez, J. (1996) *In vitro* protein-primed initiation of pneumococcal phage Cp-1 DNA replication occurs at the third 3' nucleotide of the linear template: a stepwise sliding-back mechanism. *J Mol Biol*, **260**, 369-377.
- Martin, G., Lázaro, J.M., Méndez, E. and Salas, M. (1989) Characterization of the phage  $\phi$ 29 protein p5 as a single-stranded DNA binding protein. Function in  $\phi$ 29 DNA-protein p3 replication. *Nucleic Acids Res*, **17**, 3663-3672.
- Meijer, W.J., Lewis, P.J., Errington, J. y Salas, M. (2000) Dynamic relocation of phage  $\phi$ 29 DNA during replication and the role of the viral protein p16.7. *EMBO J*, **19**, 4182-4190.
- Meijer, W.J., Horcajadas, J.A. y Salas, M. (2001a)  $\phi$ 29 family of phages. *Microbiol Mol Biol Rev*, **65**, 261-287.



- Meijer, W.J., Serna-Rico, A. y Salas, M. (2001b) Characterization of the bacteriophage  $\phi$ 29-encoded protein p16.7: a membrane protein involved in phage DNA replication. *Mol Microbiol*, **39**, 731-746.
- Mencía, M., Gella, P., Camacho, A., de Vega, M. y Salas, M. (2011) Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage  $\phi$ 29. *Proc Natl Acad Sci U S A*, **108**, 18655-18660.
- Méndez, J., Blanco, L., Esteban, J.A., Bernad, A. y Salas, M. (1992) Initiation of  $\phi$ 29 DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. *Proc Natl Acad Sci U S A*, **89**, 9579-9583.
- Méndez, J., Blanco, L. y Salas, M. (1997) Protein-primed DNA replication: a transition between two modes of priming by a unique DNA polymerase. *EMBO J*, **16**, 2519-2527.
- Merkulov, G.V. y Boeke, J.D. (1998) Libraries of green fluorescent protein fusions generated by transposition in vitro. *Gene* **222**, 213-222.
- Muñoz-Espín, D., Holguera, I., Ballesteros-Plaza, D., Carballido-López, R. y Salas, M. (2010) Viral terminal protein directs early organization of phage DNA replication at the bacterial nucleoid. *Proc Natl Acad Sci U S A*, **107**, 16548-16553.
- Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C. y Markham, A.F. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res*, **17**, 2503-2516.
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N. y Hase, T. (2000) Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res*, **28**, e63.
- Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., y Steitz, T.A. (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*, **313**, 762-766.
- Pastrana, R., Lázaro, J.M., Blanco, L., García, J.A., Méndez, E., y Salas, M. (1985) Overproduction and purification of protein p6 of *Bacillus subtilis* phage  $\phi$ 29: role in the initiation of DNA replication. *Nucleic Acids Res*, **13**, 3083-3100.

- Pečenková, T. y Pačes, V. (1999) Molecular phylogeny of  $\phi$ 29-like phages and their evolutionary relatedness to other protein-primed replicating phages and other phages hosted by gram-positive bacteria. *J Mol Evol*, **48**, 197-208.
- Peng, X., Basta, T., Haring, M., Garrett, R.A. y Prangishvili, D. (2007) Genome of the *Acidianus* bottle-shaped virus and insights into the replication and packaging mechanisms. *Virology*, **364**, 237-243.
- Peñalva, M.A. y Salas, M. (1982) Initiation of phage  $\phi$ 29 DNA replication *in vitro*: formation of a covalent complex between the terminal protein, p3, and 5'-dAMP. *Proc Natl Acad Sci U S A*, **79**, 5522-5526.
- Pérez-Arnaiz, P., Longás, E., Villar, L., Lázaro, J.M., Salas, M. y de Vega, M. (2007) Involvement of phage  $\phi$ 29 DNA polymerase and terminal protein subdomains in conferring specificity during initiation of protein-primed DNA replication. *Nucleic Acids Res*, **35**, 7061-7073.
- Pérez-Lago, L., Serrano-Heras, G., Baños, B., Lázaro, J.M., Alcorlo, M., Villar, L. y Salas, M. (2011) Characterization of *Bacillus subtilis* uracil-DNA glycosylase and its inhibition by phage  $\phi$ 29 protein p56. *Mol Microbiol*, **80**, 1657-1666.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M. Meng, E.C. y Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605–1612.
- Pinheiro, V.B., Taylor, A.I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J.C., Wengel, J., Peak-Chew, S., McLaughlin, S.H., Herdewijn, P. y Holliger, P. (2012) Synthetic genetic polymers capable of heredity and evolution. *Science*, **336**, 341-344.
- Prieto, I., Serrano, M., Lázaro, J.M., Salas, M. y Hermoso, J.M. (1988) Interaction of the bacteriophage  $\phi$ 29 protein p6 with double-stranded DNA. *Proc Natl Acad Sci U S A*, **85**, 314-318.
- Pritchard, C. y Stefano, J. (1990) Amplification detection of viral nucleic acid at subattomole levels using Q replicase. *Ann Biol Clin (Paris)*, **48**, 492-497.

- Redondo, R.A., Kupczok, A., Stift, G., y Bollback, J.P. (2013) Complete genome sequence of the novel phage MG-B1 infecting *Bacillus weihenstephanensis*. *Genome announc*, **1**, 216-13.
- Redrejo-Rodríguez, M., Muñoz-Espín, D., Holguera, I., Mencía, M. y Salas, M. (2012) Functional eukaryotic nuclear localization signals are widespread in terminal proteins of bacteriophages. *Proc Natl Acad Sci U S A*, **109**, 18482-18487.
- Redrejo-Rodríguez, M., Muñoz-Espín, D., Holguera, I., Mencía, M., y Salas, M. (2013) Nuclear and nucleoid localization are independently conserved functions in bacteriophage terminal proteins. *Mol microbiol*, **90**, 858-868.
- Redrejo-Rodríguez, M., y Salas, M. (2014) Multiple roles of genome-attached bacteriophage terminal proteins. *Virology*, **468**, 322-329.
- Reilly, B. E., y Spizizen, J. (1965) Bacteriophage deoxyribonucleate infection of competent *Bacillus subtilis*. *Journal of bacteriology*, **89**, 782-790.
- Reznikoff, W.S., Bhasin, A., Davies, D.R., Goryshin, I.Y., Mahnke, L.A., Naumann, T., Rayment, I., Steiniger-White, M. y Twining, S.S. (1999) *Tn 5*: A molecular window on transposition. *Biochem Biophys Res Commun*. **266**, 729-734.
- Rodríguez, I., Lázaro, J.M., Blanco, L., Kamtekar, S., Berman, A.J., Wang, J., Steitz, T.A., Salas, M. y de Vega, M. (2005) A specific subdomain in  $\phi 29$  DNA polymerase confers both processivity and strand-displacement capacity. *Proc Natl Acad Sci U S A*, **102**, 6407-6412.
- Ross-Macdonald, P., Sheehan, A., Roeder, G.S. y Snyder, M. (1997) A multipurpose transposon system for analyzing protein production, localization, and function in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **94**, 190-195.
- Roy, A., Kucukural, A. y Zhang, A. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, **5**, 725-738.

- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., y Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**, 1350-1354.
- Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. y Erlich, H. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.
- Salas, M., Mellado, R.P. y Viñuela, E. (1978) Characterization of a protein covalently linked to the 5' termini of the DNA of *Bacillus subtilis* phage  $\phi$ 29. *J Mol Biol*, **119**, 269-291.
- Salas, M. (1991) Protein-priming of DNA replication. *Annu Rev Biochem*, **60**, 39-71.
- Salas, M. (1999) Mechanisms of initiation of linear DNA replication in prokaryotes. *Genet Eng (NY)*, **21**, 159-171.
- Salas, M. y de Vega, M. (2006) Bacteriophage protein-primed DNA replication. En Hefferon, K. (Ed.) *Recent Advances in DNA Virus Replication*, 1 ed., Research SignPost, Ithaca, NY, 259-288.
- Salas, M. y de Vega, M. (2008) En Mahy, B. W. J. and van Regenmortel, M. H. V. (Eds.) *Encyclopedia of Virology*, 3rd edition. Oxford: Elsevier, vol. **4**, 339-406.
- Saunders, G.C., Juliet, D., Helen, C.P., y Johanne, H.C. (2001) Interlaboratory study on thermal cycler performance in controlled PCR and random amplified polymorphic DNA analyses. *Clin Chem*, **47**, 47-55.
- Sergeeva, A., Kolonin, M. G., Molldrem, J. J., Pasqualini, R., y Arap, W. (2006) Display technologies: application for the discovery of drug and gene delivery agents. *Advanced drug delivery reviews*, **58**, 1622-1654.
- Serna-Rico, A., Salas, M. y Meijer, W.J. (2002) The *Bacillus subtilis* phage  $\phi$ 29 protein p16.7, involved in  $\phi$ 29 DNA replication, is a membrane-localized single-stranded DNA-binding protein. *J Biol Chem*, **277**, 6733-6742.
- Serna-Rico, A., Muñoz-Espín, D., Villar, L., Salas, M. y Meijer, W.J. (2003) The integral membrane protein p16.7 organizes *in vivo*  $\phi$ 29 DNA replication through interaction with both the terminal protein and ssDNA. *EMBO J*, **22**, 2297-2306.

- Serrano-Heras, G., Bravo, A. y Salas, M. (2008) Phage  $\phi$ 29 protein p56 prevents viral DNA replication impairment caused by uracil excision activity of uracil-DNA glycosylase. *Proc Natl Acad Sci U S A*, **105**, 19044-19049.
- Serrano, M., Gutiérrez, J., Prieto, I., Hermoso, J.M. y Salas, M. (1989) Signals at the bacteriophage  $\phi$ 29 DNA replication origins required for protein p6 binding and activity. *EMBO J.*, **8**, 1879–1885.
- Serrano, M., Gutiérrez, C., Salas, M. and Hermoso, J.M. (1993) Superhelical path of the DNA in the nucleoprotein complex that activates the initiation of phage  $\phi$ 29 DNA replication. *J Mol Biol*, **230**, 248-259.
- Serrano, M., Gutiérrez, C., Freire, R., Bravo, A., Salas, M. y Hermoso, J.M. (1994) Phage  $\phi$ 29 protein p6: a viral histone-like protein. *Biochimie*, **76**, 981-991.
- Sheridan, D.L., Berlot, C.H., Robert, A., Inglis, F.M., Jakobsdottir, K.B., Howe J.R. y Hughes, T.E. (2002) A new way to rapidly create functional, fluorescent fusion proteins: random insertion of GFP with an in vitro transposition reaction. *BMC Neurosci* **3**: 7.
- Shiffman, D., y Cohen, S.N. (1992) Reconstruction of a *Streptomyces* linear replicon from separately cloned DNA fragments: existence of a cryptic origin of circular replication within the linear plasmid. *Proc Natl Acad Sci U S A*, **89**, 6129-6133.
- Silander, K. y Saarela, J. (2008) Whole genome amplification with  $\phi$ 29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. *Methods Mol Biol*, **439**, 1-18.
- Skerra, A. (1992) Phosphorothioate primers improve the amplification of DNA sequences by DNA polymerases with proofreading activity. *Nucleic Acids Res*, **20**, 3551-3554.
- Smith, H.O. y Wilcox, K.W. (1970) A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J Mol Biol*, **51**, 379–391.
- Snounou, G., Viriyakosol, S., Zhu, X.P., Jarra, W., Pinheiro, L., do Rosario, V.E., Thaithong, S. y Brown, K.N. (1993) High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol biochem parasitol*, **61**, 315-320.

- Soengas, M.S., Gutiérrez, C. y Salas, M. (1995) Helix-destabilizing activity of  $\phi 29$  single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication. *J Mol Biol*, **253**, 517-529.
- Sogo, J.M., García, J.A., Peñalva, M.A. y Salas, M. (1982) Structure of protein-containing replicative intermediates of *Bacillus subtilis* phage  $\phi 29$  DNA. *Virology*, **116**, 1-18.
- Sogo, J.M., Inciarte, M.R., Corral, J., Viñuela, E. y Salas, M. (1979) RNA polymerase binding sites and transcription map of the DNA of *Bacillus subtilis* phage  $\phi 29$ . *J Mol Biol*, **127**, 411-436.
- Steiner, M., Lubitz, W. y Blasi, U. (1993) The missing link in phage lysis of gram-positive bacteria: gene 14 of *Bacillus subtilis* phage  $\phi 29$  encodes the functional homolog of lambda S protein. *J Bacteriol*, **175**, 1038-1042.
- Tabor, S., y Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci U S A*, **82**, 1074-1078.
- Tawfik, D.S. y Griffiths, A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nature biotechnol*, **16**, 652-656.
- Tang, Y.W. y Persing, D.H. (2009) Diagnostic microbiology. En Schaechter, M. (Ed.) *Encyclopaedia of Microbiology*, Oxford: Elsevier, 308-320.
- Temperley, S.M., y Hay, R.T. (1992) Recognition of the adenovirus type 2 origin of DNA replication by the virally encoded DNA polymerase and preterminal proteins. *EMBO J*, **11**, 761.
- Tenover, F.C., Robert D.A., Goering, R.V. y the Molecular Typing Group of the Society for Healthcare Epidemiology of America (1997) How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. *Infect Control Hosp Epidemiol*, **18**, 426-439.
- Traven, A., Jelacic, B. y Sopta, M. (2006) Yeast Gal4: a transcriptional paradigm revisited. *EMBO reports*, **7**, 496-499.

- Twort, F.W. (1915) An investigation on the nature of ultra-microscopic viruses. *The Lancet*, **2**, 1241-1243.
- Verdun, R.E. y Karlseder, J. (2007) Replication and protection of telomeres. *Nature*, **447**, 924-931.
- Viguera, E., Canceill, D., y Ehrlich, S.D. (2001) Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J*, **20**, 2587-2595.
- Vincent, M., Xu, Y., y Kong, H. (2004) Helicase-dependent isothermal DNA amplification. *EMBO rep*, **5**, 795-800.
- Vlček, Č., y Pačes, V. (1986) Nucleotide sequence of the late region of *Bacillus* phage  $\phi$ 29 completes the 19285-bp sequence of  $\phi$ 29 genome. Comparison with the homologous sequence of phage PZA. *Gene*, **46**(2), 215-225.
- Walker, G.T. (1993) Empirical aspects of strand displacement amplification. *PCR Methods Appl*, **3**, 1-6.
- Walker, G.T., Fraiser, M.S., Schram, J.L., Little, M.C., Nadeau, J.G., y Malinowski, D.P. (1992) Strand displacement amplification-an isothermal, *in vitro* DNA amplification technique. *Nucleic Acids Res*, **20**, 1691-1696.
- Wang, S. y Hazelrigg, T. (1994) Implications for bcd mRNA localization from spatial distribution of exu protein in *Drosophila* oogenesis. *Nature* , **369**, 400-403.
- Watson, J. D. y Crick, F. H. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
- Wilkins, M. H. F., Stokes, A. R. y Wilson, H. R. (1953) Molecular structure of deoxypentose nucleic acids. *Nature*, **171**, 738-740.
- Wu, D.Y., y Wallace, R.B. (1989) The ligation amplification reaction (LAR) amplification of specific DNA sequences using sequential rounds of template-dependent ligation. *Genomics*, **4**, 560-569.
- Yang, C.C., Huang, C.H., Li, C.Y., Tsay, Y.G., Lee, S.C., y Chen, C.W. (2002) The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins. *Mol Microbiol*, **43**, 297-305.

- Yoo, S. K., y Ito, J. (1991). Sequence requirements for protein-primed DNA replication of bacteriophage PRD1. *J Mol Biol*, **218**, 779-789.
- Yoshikawa, H., Friedmann, T., y Ito, J. (1981) Nucleotide sequences at the termini of  $\phi$ 29 DNA. *Proc Natl Acad Sci U S A*, **78**, 1336-1340.
- Zaballos, A. y Salas, M. (1989) Functional domains in the bacteriophage  $\phi$ 29 terminal protein for interaction with the  $\phi$ 29 DNA polymerase and with DNA. *Nucleic Acids Res*, **17**, 10353-10366.
- Zhang, C., Lee, C. S., y Guo, P. (1994) The proximate 5' and 3' ends of the 120-base viral RNA (pRNA) are crucial for the packaging of bacteriophage  $\phi$ 29 DNA. *Virology*, **201**, 77-85.

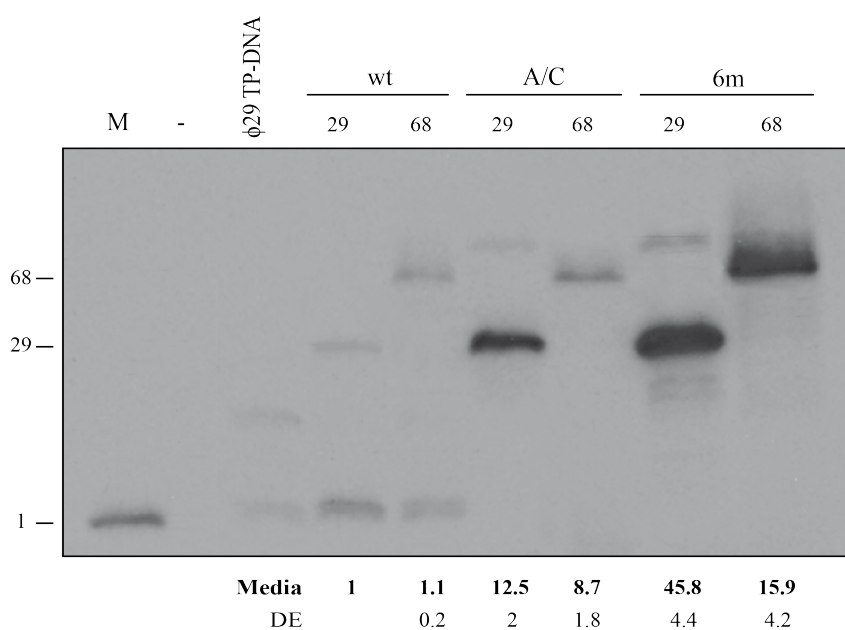


**Material Suplementario**



Nombre	Secuencia (5'-3')
MCS1	CTCGAGTGC GGCCGCAAGCTTGCATGCAGGCCTCTGCAGTCGAC
MCS2	GGTACCGAGCTCGAATTCGGATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATATG

**Tabla S1. Secuencias de los sitios de policlonaje del plásmido pETORPHI.** Se muestra en la tabla superior la secuencia de los sitios de policlonaje 1 y 2 (MCS1 y MCS2, respectivamente) presentes en el plásmido pETORPHI. La secuencia del sitio de policlonaje 1 (MCS1) contiene las dianas únicas para los enzimas de restricción XhoI, NotI, PstI y SalI. La secuencia del sitio de policlonaje 2 (MCS2) contiene las dianas únicas para los enzimas de restricción KpnI, SacI, EcoRI, NheI y NdeI.



**Figura S1. Comparación de la actividad replicativa entre las versiones de 29 y 68 pb de los orígenes wt, A/C y 6m.** La construcción de los orígenes de dsDNA y la reacción de replicación se realizó y procesó como se describe en *Materiales y Métodos* empleando [ $\alpha^{32}$ P]dATP como nucleótido marcado. Se indican las posiciones correspondientes a la iniciación [+1] (TP-dAMP) y longitud completa de las dos versiones [+29, +68] (TP-29 o 68 nucleótidos ssDNA). Carril M, marcador de iniciación (empleando el TP-DNA de  $\phi$ 29 con dATP como único nucleótido); carril  $\phi$ 29 TP-DNA, control de replicación (empleando TP-DNA de  $\phi$ 29 como sustrato y todos los nucleótidos), se observan las bandas de iniciación [+1], de elongación parcial [+15, +16] y de longitud completa [no mostrada]. El número de nucleótidos marcados incorporados por molécula de DNA replicada se tuvo en cuenta para la cuantificación (los moldes 68-mer incorporan 1.9-2.2 más moléculas de dAMP que los 29-mer, dependiendo del origen). En la parte inferior se muestran las medias con sus correspondientes desviaciones estándar de los valores obtenidos de la cuantificación de las intensidades de las bandas correspondientes al producto de longitud total. Los valores se normalizaron con los valores obtenidos para el origen wt 29-mer (wt = 1). Los valores son la media de, al menos, tres experimentos independientes.

